



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments

**Citation for published version:**

He, D, Hughes, S, Vanden-hehir, S, Georgiev, A, Altenbach, K, Tarrant, E, Mackay, CL, Waldron, KJ, Clarke, DJ & Marles-wright, J 2016, 'Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments', *eLIFE*, vol. 5, e18972. <https://doi.org/10.7554/eLife.18972>

**Digital Object Identifier (DOI):**

[10.7554/eLife.18972](https://doi.org/10.7554/eLife.18972)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

eLIFE

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# Structural characterization of encapsulated ferritin provides insight into iron storage in bacterial nanocompartments

Didi He<sup>1</sup>, Sam Hughes<sup>2</sup>, Sally Vanden-Hehir<sup>2</sup>, Atanas Georgiev<sup>1</sup>, Kirsten Altenbach<sup>1</sup>, Emma Tarrant<sup>3</sup>, C Logan Mackay<sup>2</sup>, Kevin J Waldron<sup>3</sup>, David J Clarke<sup>2\*</sup>, Jon Marles-Wright<sup>1,4\*</sup>

<sup>1</sup>Institute of Quantitative Biology, Biochemistry and Biotechnology, School of Biological Sciences, The University of Edinburgh, Edinburgh, United Kingdom; <sup>2</sup>The School of Chemistry, The University of Edinburgh, Edinburgh, United Kingdom; <sup>3</sup>Institute for Cell and Molecular Biosciences, Newcastle University, Newcastle upon Tyne, United Kingdom; <sup>4</sup>School of Biology, Newcastle University, Newcastle upon Tyne, United Kingdom

**Abstract** Ferritins are ubiquitous proteins that oxidise and store iron within a protein shell to protect cells from oxidative damage. We have characterized the structure and function of a new member of the ferritin superfamily that is sequestered within an encapsulin capsid. We show that this encapsulated ferritin (EncFtn) has two main alpha helices, which assemble in a metal dependent manner to form a ferroxidase center at a dimer interface. EncFtn adopts an open decameric structure that is topologically distinct from other ferritins. While EncFtn acts as a ferroxidase, it cannot mineralize iron. Conversely, the encapsulin shell associates with iron, but is not enzymatically active, and we demonstrate that EncFtn must be housed within the encapsulin for iron storage. This encapsulin nanocompartment is widely distributed in bacteria and archaea and represents a distinct class of iron storage system, where the oxidation and mineralization of iron are distributed between two proteins.

DOI: [10.7554/eLife.18972.001](https://doi.org/10.7554/eLife.18972.001)

\*For correspondence: dave.clarke@ed.ac.uk (DJC); jon.marles-wright@ncl.ac.uk (JM-W)

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 28

**Received:** 24 June 2016

**Accepted:** 14 August 2016

**Published:** 16 August 2016

**Reviewing editor:** Richard Losick, Harvard University, United States

© Copyright He et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

## Introduction

Encapsulin nanocompartments are a family of proteinaceous metabolic compartments that are widely distributed in bacteria and archaea (Sutter et al., 2008; Akita et al., 2007; McHugh et al., 2014; Contreras et al., 2014). They share a common architecture, comprising an icosahedral shell formed by the oligomeric assembly of a protein, encapsulin, that is structurally related to the HK97 bacteriophage capsid protein gp5 (Helgstrand et al., 2003). Gp5 is known to assemble as a 66 nm diameter icosahedral shell of 420 subunits. In contrast, both the *Pyrococcus furiosus* (Akita et al., 2007) and *Myxococcus xanthus* (McHugh et al., 2014) encapsulin shell-proteins form 32 nm icosahedra with 180 subunits; while the *Thermotoga maritima* (Sutter et al., 2008) encapsulin is smaller still with a 25 nm, 60-subunit icosahedron. The high structural similarity of the encapsulin shell-proteins to gp5 suggests a common evolutionary origin for these proteins (McHugh et al., 2014).

The genes encoding encapsulin proteins are found downstream of genes for dye-dependent peroxidase (DyP) family enzymes (Roberts et al., 2011), or encapsulin-associated ferritins (EncFtn) (He and Marles-Wright, 2015). Enzymes in the DyP family are active against polyphenolic

**eLife digest** Iron is essential for life as it is a key component of many different enzymes that participate in processes such as energy production and metabolism. However, iron can also be highly toxic to cells because it readily reacts with oxygen. This reaction can damage DNA, proteins and the membranes that surround cells.

To balance the cell's need for iron against its potential damaging effects, organisms have evolved iron storage proteins known as ferritins that form cage-like structures. The ferritins convert iron into a less reactive form that is mineralised and safely stored in the central cavity of the ferritin cage and is available for cells when they need it.

Recently, a new family of ferritins known as encapsulated ferritins have been found in some microorganisms. These ferritins are found in bacterial genomes with a gene that codes for a protein cage called an encapsulin. Although the structure of the encapsulin cage is known to look like the shell of a virus, the structure that the encapsulated ferritin itself forms is not known. It is also not clear how encapsulin and the encapsulated ferritin work together to store iron.

He et al. have now used the techniques of X-ray crystallography and mass spectrometry to determine the structure of the encapsulated ferritin found in some bacteria. The encapsulated ferritin forms a ring-shaped doughnut in which ten subunits of ferritin are arranged in a ring; this is totally different from the enclosed cages that other ferritins form.

Biochemical studies revealed that the encapsulated ferritin is able to convert iron into a less reactive form, but it cannot store iron on its own since it does not form a cage. Thus, the encapsulated ferritin needs to be housed within the encapsulin cage to store iron.

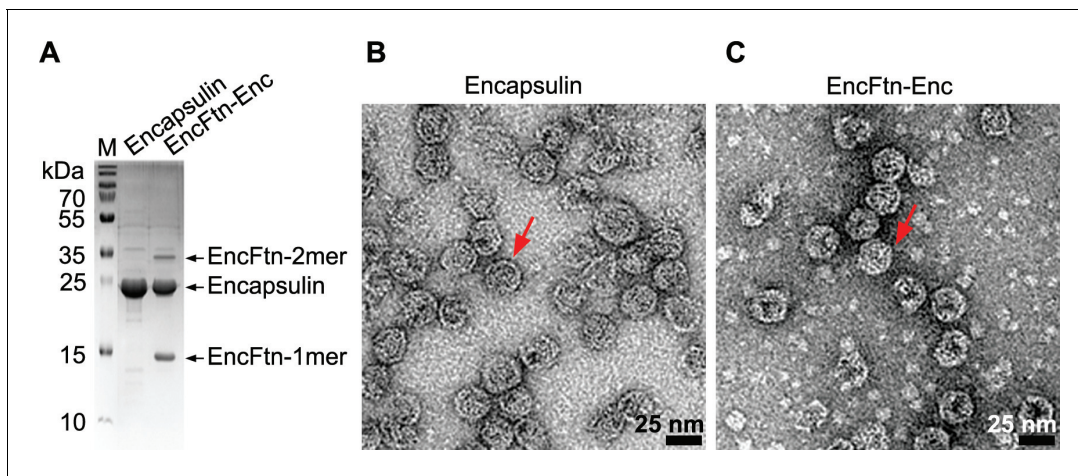
Further work is needed to investigate how iron moves into the encapsulin cage to reach the ferritin proteins. Some organisms have both standard ferritin cages and encapsulated ferritins; why this is the case also remains to be discovered.

DOI: [10.7554/eLife.18972.002](https://doi.org/10.7554/eLife.18972.002)

compounds such as azo dyes and lignin breakdown products; although their physiological function and natural substrates are not known (Roberts et al., 2011). Ferritin family proteins are found in all kingdoms and have a wide range of activities, including ribonucleotide reductase (Aberg et al., 1993), protecting DNA from oxidative damage (Grant et al., 1998), and iron storage (Bradley et al., 2014). The classical iron storage ferritin nanocages are found in all kingdoms and are essential in eukaryotes; they play a central role in iron homeostasis, where they protect the cell from toxic free  $\text{Fe}^{2+}$  by oxidizing it and storing the resulting  $\text{Fe}^{3+}$  as ferrihydrite minerals within their central cavity.

The encapsulin-associated enzymes are sequestered within the icosahedral shell through interactions between the shell's inner surface and a short localization sequence (Gly-Ser-Leu-Lys) appended to their C-termini (Sutter et al., 2008). This motif is well-conserved, and the addition of this sequence to heterologous proteins is sufficient to direct them to the interior of encapsulins (Rurup et al., 2014; 2015; Cassidy-Amstutz et al., 2016).

A recent study of the *Myxococcus xanthus* encapsulin showed that it sequesters a number of different EncFtn proteins and acts as an 'iron-megastore' to protect these bacteria from oxidative stress (McHugh et al., 2014). At 32 nm in diameter, it is much larger than other members of the ferritin superfamily, such as the 12 nm 24-subunit classical ferritin nanocage and the 8 nm 12-subunit Dps (DNA-binding protein from starved cells) complex (Grant et al., 1998; Andrews, 2010); and is thus capable of sequestering up to ten times more iron than these ferritins (McHugh et al., 2014). The primary sequences of EncFtn proteins have Glu-X-X-His metal coordination sites, which are shared features of the ferritin family proteins (Andrews, 2010). Secondary structure prediction identifies two major  $\alpha$ -helical regions in these proteins; this is in contrast to other members of the ferritin superfamily, which have four major  $\alpha$ -helices (Supplementary file 1). The 'half-ferritin' primary sequence of the EncFtn family and their association with encapsulin nanocompartments suggests a distinct biochemical and structural organization to other ferritin family proteins. The *Rhodospirillum rubrum* EncFtn protein (Rru\_A0973) shares 33% protein sequence identity with the *M. xanthus* (MXAN\_4464), 53% with the *T. maritima* (Tmari\_0787), and 29% with the *P. furiosus*



**Figure 1.** Purification of recombinant *R. rubrum* encapsulin nanocompartments. (A) Recombinantly expressed encapsulin (Enc) and co-expressed EncFtn-Enc were purified by sucrose gradient ultracentrifugation from *E. coli* B834(DE3) grown in SeMet medium. Samples were resolved by 18% acrylamide SDS-PAGE; the position of the proteins found in the complexes as resolved on the gel are shown with arrows. (B/C) Negative stain TEM image of recombinant encapsulin and EncFtn-Enc nanocompartments. Samples were imaged at 143,000 x magnification, with scale bar shown as 25 nm. Representative encapsulin and EncFtn-Enc complexes are indicated with red arrows.

DOI: [10.7554/eLife.18972.003](https://doi.org/10.7554/eLife.18972.003)

The following figure supplement is available for figure 1:

**Figure supplement 1.** Full-frame transmission electron micrographs of *R. rubrum* nanocompartments.

DOI: [10.7554/eLife.18972.004](https://doi.org/10.7554/eLife.18972.004)

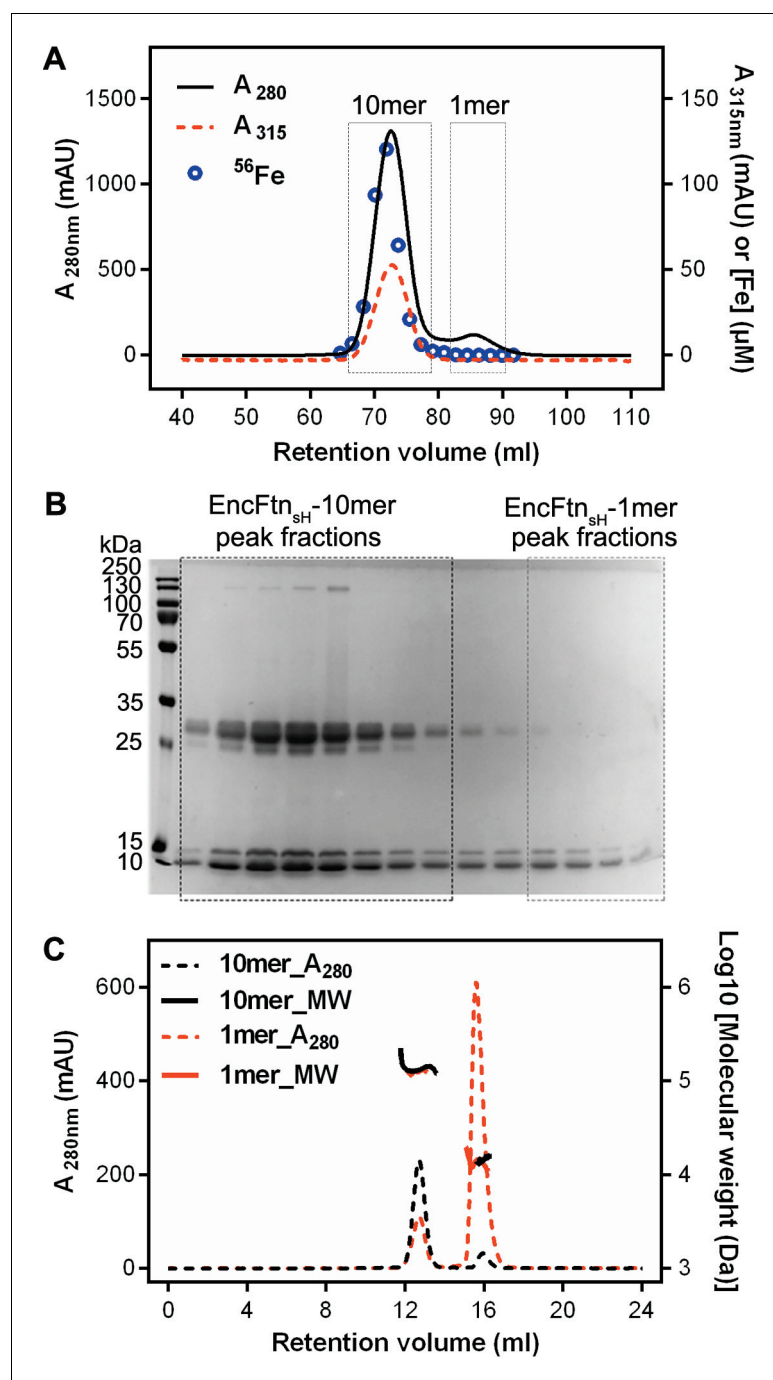
(PF1192) homologues. The GXXH motifs are strictly conserved in each of these species (**Supplementary file 1**).

Here we investigate the structure and biochemistry of EncFtn in order to understand iron storage within the encapsulin nanocompartment. We have produced recombinant encapsulin (Enc) and EncFtn from the aquatic purple-sulfur bacterium *R. rubrum*, which serves as a model organism for the study of the control of the bacterial nitrogen fixation machinery (**Pope et al., 1985**), in *Escherichia coli*. Analysis by transmission electron microscopy (TEM) indicates that their co-expression leads to the production of an icosahedral nanocompartment with encapsulated EncFtn. The crystal structure of a truncated hexahistidine-tagged variant of the EncFtn protein (EncFtn<sub>SH</sub>) shows that it forms a decameric structure with an annular ‘ring-doughnut’ topology, which is distinct from the four-helical bundles of the 24meric ferritins (**Lawson et al., 1991**) and dodecahedral DPS proteins (**Grant et al., 1998**). We identify a symmetrical iron bound ferroxidase center (FOC) formed between subunits in the decamer and additional metal-binding sites close to the center of the ring and on the outer surface. We also demonstrate the metal-dependent assembly of EncFtn decamers using native PAGE, analytical gel-filtration, and native mass spectrometry. Biochemical assays show that EncFtn is active as a ferroxidase enzyme. Through site-directed mutagenesis we show that the conserved glutamic acid and histidine residues in the FOC influence protein assembly and activity. We use our combined structural and biochemical data to propose a model for the EncFtn-catalyzed sequestration of iron within the encapsulin shell.

## Results

### Assembly of *R. rubrum* EncFtn encapsulin nanocompartments in *E. coli*

We produced recombinant *R. rubrum* encapsulin nanocompartments in *E. coli* by co-expression of the encapsulin (Rru\_A0974) and EncFtn (Rru\_A0973) proteins, and purified these by sucrose gradient ultra-centrifugation (**Figure 1A**) (**Sutter et al., 2008**). TEM imaging of uranyl acetate-stained samples revealed that, when expressed in isolation, the encapsulin protein forms empty compartments with an average diameter of 24 nm (**Figure 1B** and **Figure 1—figure supplement 1A/C**), consistent with the appearance and size of the *T. maritima* encapsulin (**Sutter et al., 2008**). We were not able



**Figure 2.** Purification of recombinant *R. rubrum* EncFtn<sub>SH</sub>. (A) Recombinant SeMet-labeled EncFtn<sub>SH</sub> produced with 1 mM Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub> in the growth medium was purified by nickel affinity chromatography and size-exclusion chromatography using a Superdex 200 16/60 column (GE Healthcare). Chromatogram traces measured at 280 nm and 315 nm are shown with the results from ICP-MS analysis of the iron content of the fractions collected during the experiment. The peak around 73 ml corresponds to a molecular weight of around 130 kDa when compared to calibration standards; this is consistent with a decamer of EncFtn<sub>SH</sub>. The small peak at 85 ml corresponds to the 13 kDa monomer compared to the standards. Only the decamer peak contains significant amounts of iron as indicated by the ICP-MS analysis. (B) Peak fractions from the gel filtration run were resolved by 15% acrylamide SDS-PAGE and stained with Coomassie blue stain. The bands around 13 kDa and 26 kDa correspond to EncFtn<sub>SH</sub>, as identified by MALDI peptide mass fingerprinting. The band at 13 kDa is consistent with the monomer mass, while the band at 26 kDa is consistent with a dimer of EncFtn<sub>SH</sub>. The dimer species only appears in the decamer fractions. (C) SEC-MALLS analysis of EncFtn<sub>SH</sub> from decamer fractions and monomer fractions allows assignment of

Figure 2 continued on next page

Figure 2 continued

an average mass of 132 kDa to decamer fractions and 13 kDa to monomer fractions, consistent with decamer and monomer species (Table 2).  
DOI: 10.7554/eLife.18972.005

to resolve any higher-order structures of EncFtn by TEM. Protein purified from co-expression of the encapsulin and EncFtn resulted in 24 nm compartments with regions in the center that exclude stain, consistent with the presence of the EncFtn within the encapsulin shell (Figure 1C and Figure 1—figure supplement 1B/C).

R. rubrum EncFtn forms a metal-ion stabilized decamer in solution

We purified recombinant *R. rubrum* EncFtn as both the full-length sequence (140 amino acids) and a truncated C-terminal hexahistidine-tagged variant (amino acids 1–96 plus the tag; herein EncFtn<sub>sH</sub>). In both cases the elution profile from size-exclusion chromatography (SEC) displayed two peaks (Figure 2A). SDS-PAGE analysis of fractions from these peaks showed that the high molecular weight peak was partially resistant to SDS and heat-induced denaturation; in contrast, the low molecular weight peak was consistent with monomeric mass of 13 kDa (Figure 2B). MALDI peptide mass fingerprinting of these bands confirmed the identity of both as EncFtn. Inductively coupled plasma mass spectrometry (ICP-MS) analysis of the SEC fractions showed 100 times more iron in the oligomeric fraction than the monomer (Figure 2A, blue scatter points; Table 1), suggesting that EncFtn oligomerization is associated with iron binding. In order to determine the iron-loading stoichiometry in the EncFtn complex, further ICP-MS experiments were performed using selenomethionine (SeMet)-labelled protein EncFtn (Table 1). In these experiments, we observed sub-stoichiometric metal binding, which is in contrast to the classical ferritins (Le Brun et al., 2010). Size-exclusion chromatography with multi-angle laser light scattering (SEC-MALLS) analysis of

**Table 1.** Determination of the Fe/EncFtn<sub>sH</sub> protein ratio by ICP-MS. EncFtn<sub>sH</sub> was purified as a SeMet derivative from *E. coli* B834(DE3) cells grown in SeMet medium with 1 mM Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>. Fractions from SEC were collected, acidified and analysed by ICP-MS. EncFtn<sub>sH</sub> concentration was calculated based on the presence of two SeMet per mature monomer. Samples where the element was undetectable are labelled with n.d. These data were collected from EncFtn<sub>sH</sub> fractions from a single gel-filtration run.

Peak	EncFtn <sub>sH</sub> retention volume (ml)	Element concentration (μM)				Derived EncFtn <sub>sH</sub> concentration (μM)	Derived Fe/EncFtn <sub>sH</sub> monomer
		Ca	Fe	Zn	Se		
Decamer	66.5	n.d.	6.7	n.d.	24.6	12.3	0.5
	68.3	n.d.	28.4	n.d.	124.5	62.3	0.5
	70.1	2.9	93.7	2.4	301.7	150.9	0.6
	71.9	6.9	120.6	3.7	379.8	189.9	0.6
	73.7	1.9	64.4	0.8	240.6	120.3	0.5
	75.5	0.9	21.1	n.d.	101.7	50.8	0.4
	77.3	n.d.	6.2	n.d.	42.6	21.3	0.3
	79.1	0.1	2.4	n.d.	26.5	13.3	0.2
	80.9	1.0	1.5	n.d.	22.3	11.2	0.1
	82.7	n.d.	0.2	n.d.	29.2	14.6	n.d.
Monomer	84.5	n.d.	0.1	n.d.	34.9	17.5	n.d.
	86.3	n.d.	n.d.	n.d.	28.9	14.4	n.d.
	88.1	n.d.	n.d.	n.d.	17.4	8.7	n.d.
	89.9	n.d.	n.d.	n.d.	5.5	2.8	n.d.
	91.7	n.d.	n.d.	n.d.	0.1	0.07	0.2

DOI: 10.7554/eLife.18972.006



**Table 2.** Estimates of EncFtn<sub>sH</sub> molecular weight from SEC-MALLS analysis. EncFtn<sub>sH</sub> was purified from *E. coli* BL21(DE3) grown in minimal medium (MM) by nickel affinity chromatography and size-exclusion chromatography. Fractions from two peaks (decamer and monomer) were pooled separately (**Figure 1C**) and analysed by SEC-MALLS using a Superdex 200 10/300 GL column (GE Healthcare) and Viscotek SEC-MALLS instruments (Malvern Instruments) (**Figure 2C**). The decamer and monomer peaks were both symmetric and monodisperse, allowing the estimation of the molecular weight of the species in these fractions (**Folta-Stogniew, 2006**). The molecular weights are quoted to the nearest kDa due to the resolution limit of the instrument. The proteins analyzed by SEC-MALLS came from single protein preparation.

Molecular Weight (kDa)	Decamer peak	Monomer peak
Theoretical	133	13
EncFtn <sub>sH</sub> -decamer fractions	132	15
EncFtn <sub>sH</sub> -monomer fractions	126	13

DOI: [10.7554/eLife.18972.007](https://doi.org/10.7554/eLife.18972.007)

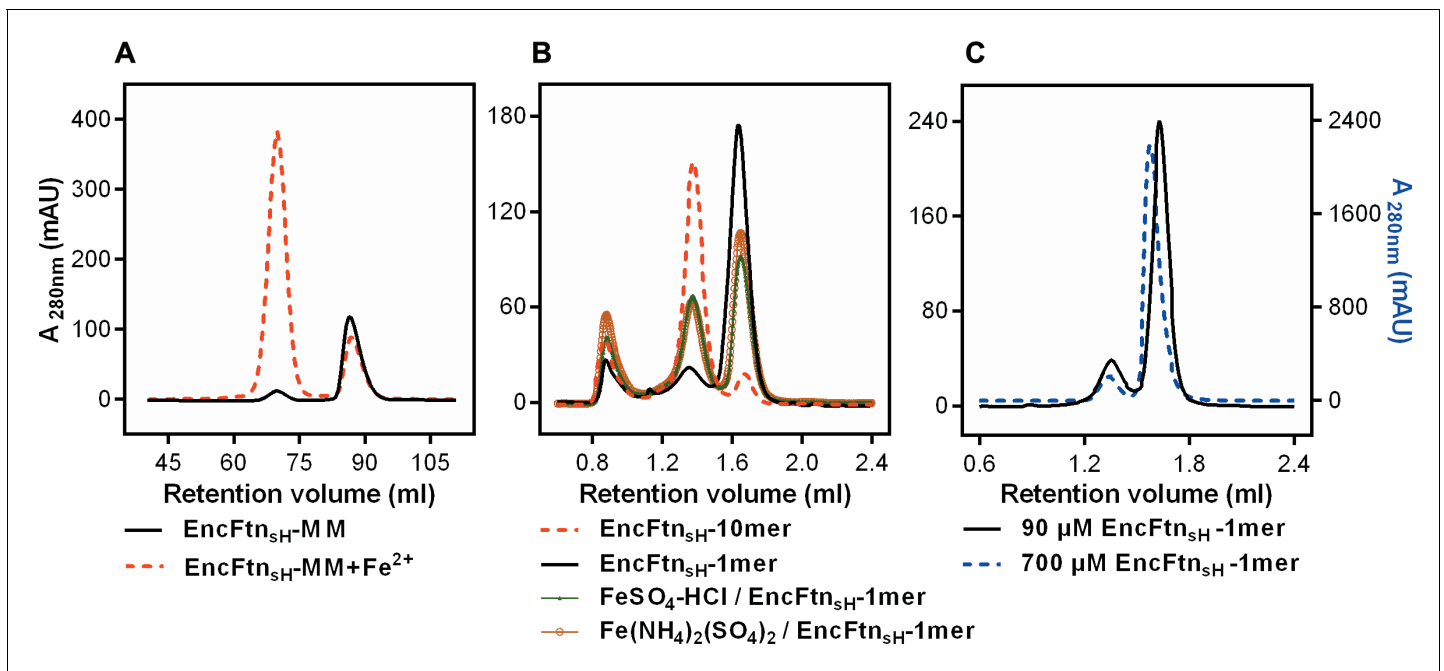
samples taken from each peak gave calculated molecular weights consistent with a decamer for the high molecular weight peak and a monomer for the low molecular weight peak (**Figure 2C, Table 2**).

We purified EncFtn<sub>sH</sub> from *E. coli* grown in MM with or without the addition of 1 mM Fe (NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>. The decamer to monomer ratio in the sample purified from cells grown in iron-supplemented media was 4.5, while that from the iron-free media was 0.11, suggesting that iron induces the oligomerization of EncFtn<sub>sH</sub> *in vivo* (**Figure 3A, Table 3**). To test the metal-dependent oligomerization of EncFtn<sub>sH</sub> *in vitro*, we incubated the protein with various metal cations and subjected samples to analytical SEC and non-denaturing PAGE. Of the metals tested, only Fe<sup>2+</sup>, Zn<sup>2+</sup> and Co<sup>2+</sup> induced the formation of significant amounts of the decamer (**Figure 3B, Figure 3—figure supplement 1/2**). While Fe<sup>2+</sup> induces the multimerization of EncFtn<sub>sH</sub>, Fe<sup>3+</sup> in the form of FeCl<sub>3</sub> does not have this effect on the protein, highlighting the apparent preference this protein has for the ferrous form of iron. To determine if the oligomerization of EncFtn<sub>sH</sub> was concentration dependent we performed analytical SEC at 90 and 700 μM protein concentration (**Figure 3C**). At the higher concentration, no increase in the decameric form of EncFtn was observed; however, the shift in the major peak from the position of the monomer species indicated a tendency to dimerize at high concentration.

## Crystal structure of EncFtn<sub>sH</sub>

We determined the crystal structure of EncFtn<sub>sH</sub> by molecular replacement to 2.0 Å resolution (see **Table 1** for X-ray data collection and refinement statistics). The crystallographic asymmetric unit contained thirty monomers of EncFtn with visible electron density for residues 7 – 96 in each chain. The protein chains were arranged as three identical annular decamers, each with D<sub>5</sub> symmetry. The decamer has a diameter of 7 nm and thickness of 4 nm (**Figure 4A**). The monomer of EncFtn has an N-terminal 3<sub>10</sub>-helix that precedes two 4 nm long antiparallel α-helices arranged with their long axes at 25° to each other; these helices are followed by a shorter 1.4 nm helix projecting at 70° from α2 (**Figure 4B**). The C-terminal region of the crystallized construct extends from the outer circumference of the ring, indicating that the encapsulin localization sequence in the full-length protein is on the exterior of the ring and is thus free to interact with its binding site on the encapsulin shell protein (**Sutter et al., 2008**).

The monomer of EncFtn<sub>sH</sub> forms two distinct dimer interfaces within the decamer (**Figure 4 C/D**). The first dimer is formed from two monomers arranged antiparallel to each other, with α1 from each monomer interacting along their lengths and α3 interdigitating with α2 and α3 of the partner chain. This interface buries one third of the surface area from each partner and is stabilized by thirty hydrogen bonds and fourteen salt bridges (**Figure 4C**). The second dimer interface forms an antiparallel four-helix bundle between helices 1 and 2 from each monomer (**Figure 4D**). This interface is less extensive than the first and is stabilized by twenty-one hydrogen bonds, six salt bridges, and a number of metal ions.



**Figure 3.** Effect of  $\text{Fe}^{2+}$  and protein concentration on the oligomeric state of EncFtn<sub>sH</sub> in solution. (A) Recombinant EncFtn<sub>sH</sub> was purified by Gel filtration Superdex 200 chromatography from *E. coli* BL21(DE3) grown in MM or in MM supplemented with 1 mM  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$  (MM+ $\text{Fe}^{2+}$ ). A higher proportion of decamer (peak between 65 and 75 ml) is seen in the sample purified from MM+ $\text{Fe}^{2+}$  compared to EncFtn<sub>sH</sub>-MM, indicating that  $\text{Fe}^{2+}$  facilitates the multimerization of EncFtn<sub>sH</sub> *in vivo*. (B) EncFtn<sub>sH</sub>-monomer was incubated with one molar equivalent of  $\text{Fe}^{2+}$  salts for two hours prior to analytical gel-filtration using a Superdex 200 PC 3.2/30 column (GE Healthcare). Both  $\text{Fe}^{2+}$  salts tested induced the formation of decamer indicated by the peak between 1.2 and 1.6 ml. Monomeric and decameric samples of EncFtn<sub>sH</sub> are shown as controls. Peaks around 0.8 ml were seen as protein aggregation. (C) Analytical gel filtration of EncFtn monomer at different concentrations to illustrate the effect of protein concentration on multimerization. The major peak shows a shift towards a dimer species at high concentration of protein, but the ratio of this peak (1.5–1.8 ml) to the decamer peak (1.2–1.5 ml) does not change when compared to the low concentration sample.

DOI: [10.7554/eLife.18972.008](https://doi.org/10.7554/eLife.18972.008)

The following figure supplements are available for figure 3:

**Figure supplement 1.** Effect of metal ions on the oligomeric state of EncFtn<sub>sH</sub> in solution.

DOI: [10.7554/eLife.18972.009](https://doi.org/10.7554/eLife.18972.009)

**Figure supplement 2.** PAGE analysis of the effect of metal ions on the oligomeric state of EncFtn<sub>sH</sub>.

DOI: [10.7554/eLife.18972.010](https://doi.org/10.7554/eLife.18972.010)

The arrangement of ten monomers in alternating orientation forms the decamer of EncFtn, which assembles as a pentamer of dimers (Figure 4A). Each monomer lies at 45° relative to the vertical central-axis of the ring, with the N-termini of alternating subunits capping the center of the ring at each end, while the C-termini are arranged around the circumference. The central hole in the ring is 2.5 nm at its widest in the center of the complex, and 1.5 nm at its narrowest point near the outer surface, although it should be noted that a number of residues at the N-terminus are not visible in the crystallographic electron density and these may occupy the central channel. The surface of the decamer has distinct negatively charged patches, both within the central hole and on the outer circumference, which form spokes through the radius of the complex (Figure 4—figure supplement 1).

### EncFtn ferroxidase center

The electron density maps of the initial EncFtn<sub>sH</sub> model displayed significant positive peaks in the mFo-DFc map at the center of the 4-helix bundle dimer (Figure 5—figure supplement 1). Informed by the ICP-MS data indicating the presence of iron in the protein we collected diffraction data at the experimentally determined iron absorption edge (1.74 Å) and calculated an anomalous difference Fourier map using this data. Inspection of this map showed two 10-sigma peaks between residues Glu32, Glu62 and His65 of two adjacent chains, and a statistically smaller 5-sigma peak between



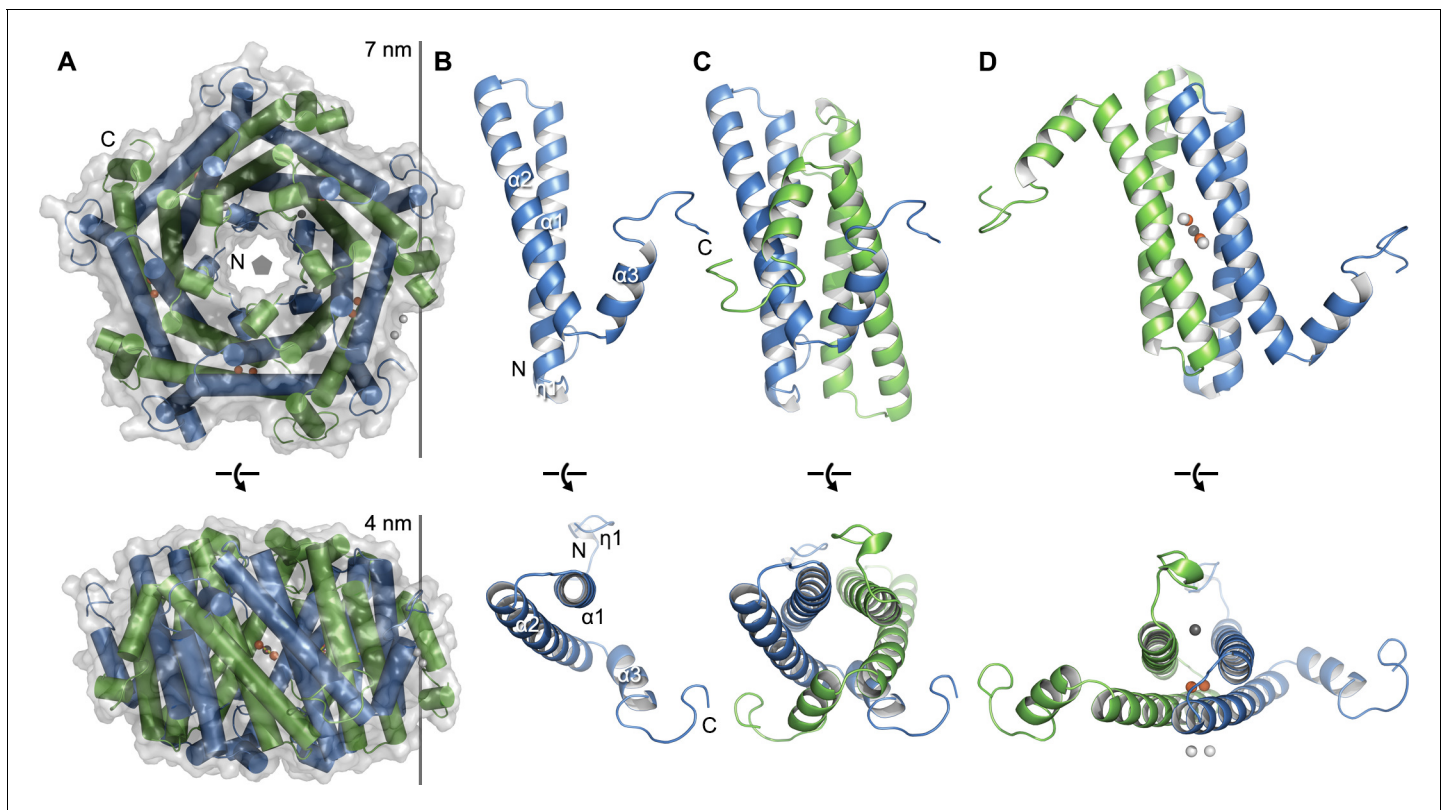
**Table 3.** Gel-filtration peak area ratios for EncFtn<sub>sH</sub> decamer and monomer on addition of different metal ions. EncFtn<sub>sH</sub> was produced in *E. coli* BL21(DE3) cultured in MM and MM with 1 mM Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub> (MM+Fe<sup>2+</sup>) and purified by gel-filtration chromatography using an Superdex 200 16/60 column (GE Healthcare). Monomer fractions of EncFtn<sub>sH</sub> purified from MM were pooled and run in subsequent analytical gel-filtration runs over the course of three days. Samples of EncFtn<sub>sH</sub> monomer were incubated with one molar equivalent of metal ion salts at room temperature for two hours before analysis by analytical gel filtration chromatography (AGF) using a Superdex 200 10/300 GL column. The area for resulting protein peaks were calculated using the Unicorn software (GE Healthcare); peak ratios were calculated to quantify the propensity of EncFtn<sub>sH</sub> to multimerize in the presence of the different metal ions. The change in the ratios of monomer to decamer over the three days of experiments may be a consequence of experimental variability, or the propensity of this protein to equilibrate towards decamer over time. The increased decamer: monomer ratio seen in the presence of Fe<sup>2+</sup>, Co<sup>2+</sup>, and Zn<sup>2+</sup> indicates that these metal ions facilitate multimerization of the EncFtn<sub>sH</sub> protein, while the other metal ions tested do not appear to induce multimerization. The analytical gel filtration experiment was repeated twice using two independent preparations of protein, of which values calculated from one sample are presented here.

Method	Sample	Monomer area	Decamer area	Decamer/Monomer
Gel filtration Superdex 200 chromatography	EncFtn <sub>sH</sub> -MM	64.3	583.6	0.1
	EncFtn <sub>sH</sub> -MM+Fe <sup>2+</sup>	1938.4	426.4	4.5
Analytical Gel filtration Day1	EncFtn <sub>sH</sub> -decamer fractions	20.2	1.8	11.2
	EncFtn <sub>sH</sub> -monomer fractions	2.9	21.9	0.1
	Fe(NH <sub>4</sub> ) <sub>2</sub> (SO <sub>4</sub> ) <sub>2</sub> /EncFtn <sub>sH</sub> -monomer	11.0	13.0	0.8
	FeSO <sub>4</sub> -HCl/EncFtn <sub>sH</sub> -monomer	11.3	11.4	1.0
Analytical Gel filtration Day2	EncFtn <sub>sH</sub> -monomer fractions	8.3	22.8	0.4
	CoCl <sub>2</sub> /EncFtn <sub>sH</sub> -monomer	17.7	14.5	1.2
	MnCl <sub>2</sub> /EncFtn <sub>sH</sub> -monomer	3.1	30.5	0.1
	ZnSO <sub>4</sub> /EncFtn <sub>sH</sub> -monomer	20.4	9.0	2.3
	FeCl <sub>3</sub> /EncFtn <sub>sH</sub> -monomer	3.9	28.6	0.1
Analytical Gel filtration Day3	EncFtn <sub>sH</sub> -monomer fractions	6.3	23.4	0.3
	MgSO <sub>4</sub> /EncFtn <sub>sH</sub> -monomer	5.8	30.2	0.2
	Ca acetate/EncFtn <sub>sH</sub> -monomer	5.6	25.2	0.2

DOI: 10.7554/eLife.18972.011

residues Glu31 and Glu34 of the two chains. Modeling metal ions into these peaks and refinement of the anomalous scattering parameters allowed us to identify these as two iron ions and a calcium ion respectively (**Figure 5A**). An additional region of asymmetric electron density near the di-iron binding site in the mFo-DFc map was modeled as glycolic acid, presumably a breakdown product of the PEG 3350 used for crystallization. This di-iron center has an Fe-Fe distance of 3.5 Å, Fe-Glu-O distances between 2.3 and 2.5 Å, and Fe-His-N distances of 2.5 Å (**Figure 5B**). This coordination geometry is consistent with the di-nuclear ferroxidase center (FOC) found in ferritin (**Bertini et al., 2012**). It is interesting to note that although we did not add any additional iron to the crystallization trials, the FOC was fully occupied with iron in the final structure, implying that this site has a very high affinity for iron.

The calcium ion coordinated by Glu31 and Glu34 adopts heptacoordinate geometry, with coordination distances of 2.5 Å between the metal ion and carboxylate oxygens of Glu31 and Glu34 (E31/34-site). A number of ordered solvent molecules are also coordinated to this metal ion at a distance of 2.5 Å. This heptacoordinate geometry is common in crystal structures with calcium ions (**Figure 5C**) (**Katz et al., 1996**). While ICP-MS indicated that there were negligible amounts of calcium in the purified protein, the presence of 140 mM calcium acetate in the crystallization mother liquor favors the coordination of calcium at this site. The fact that the protein does not multimerize in solution in the presence of Fe<sup>3+</sup> may indicate that these metal binding sites have a lower affinity for the ferric form of iron, which is the product of the ferroxidase reaction. A number of additional metal-ions were present at the outer circumference of at least one decamer in the asymmetric unit (**Figure 5D**). These ions are coordinated by His57, Glu61 and Glu64 from both chains in the FOC



**Figure 4.** Crystal structure of EncFtn<sub>SH</sub>. (A) Overall architecture of EncFtn<sub>SH</sub>. Transparent solvent accessible surface view with  $\alpha$ -helices shown as tubes and bound metal ions as spheres. Alternating subunits are colored blue and green for clarity. The doughnut-like decamer is 7 nm in diameter and 4.5 nm thick. (B) Monomer of EncFtn<sub>SH</sub> shown as a secondary structure cartoon. (C/D) Dimer interfaces formed in the decameric ring of EncFtn<sub>SH</sub>. Subunits are shown as secondary structure cartoons and colored blue and green for clarity. Bound metal ions are shown as orange spheres for Fe<sup>3+</sup> and grey and white spheres for Ca<sup>2+</sup>.

DOI: [10.7554/eLife.18972.012](https://doi.org/10.7554/eLife.18972.012)

The following figure supplement is available for figure 4:

**Figure supplement 1.** Electrostatic surface of EncFtn<sub>SH</sub>.

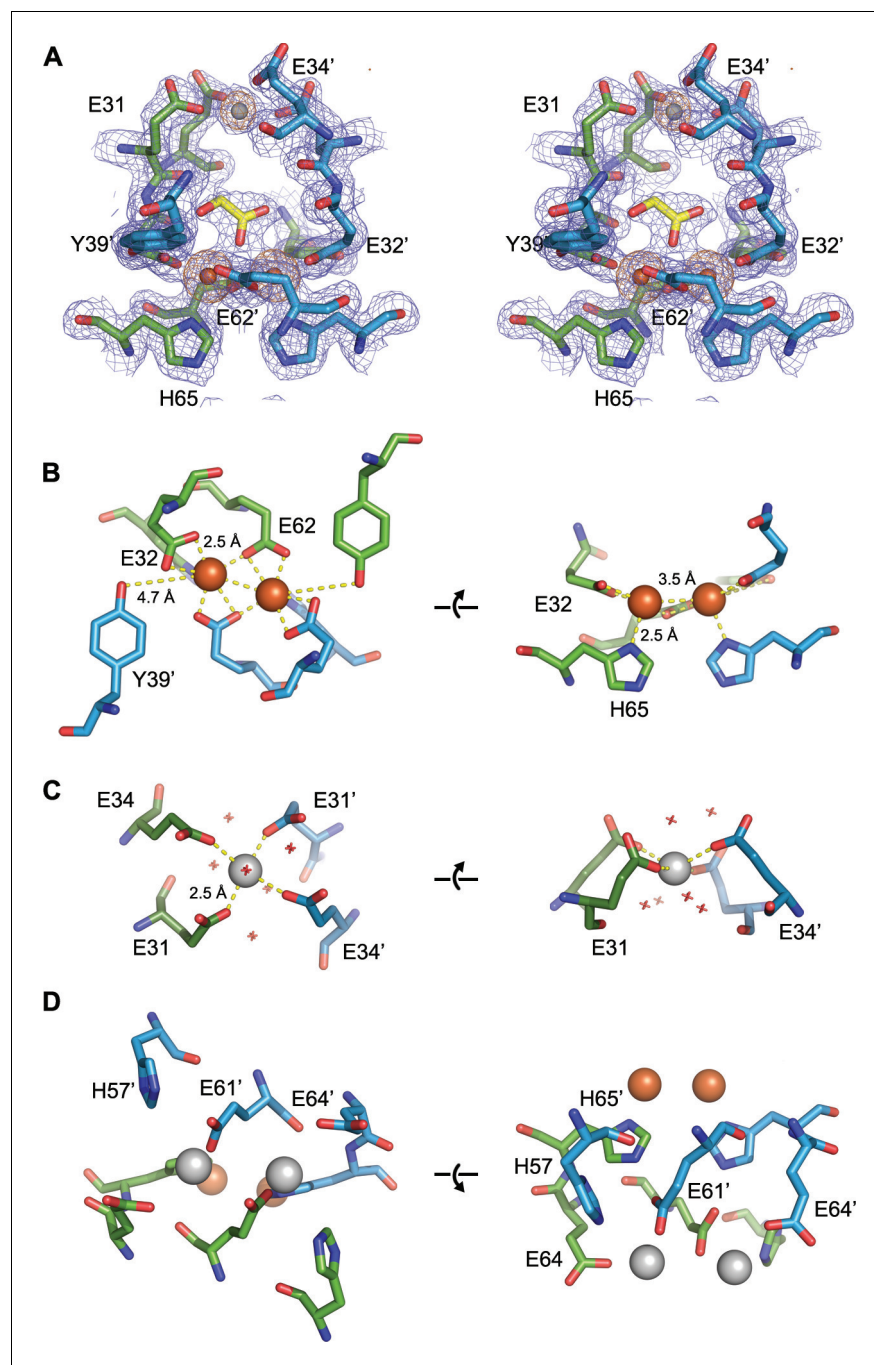
DOI: [10.7554/eLife.18972.013](https://doi.org/10.7554/eLife.18972.013)

dimer and are 4.5 Å apart; Fe-Glu-O distances are between 2.5 and 3.5 Å and the Fe-His-N distances are 4 and 4.5 Å.

Structural alignment of the di-iron binding site of EncFtn<sub>SH</sub> to the FOC of *Pseudo-nitzschia multiseries* ferritin (PmFtn, PDB ID: 4ITW) reveals a striking similarity between the metal binding sites of EncFtn<sub>SH</sub> and the classical ferritins (Pfaffen et al., 2013) (Figure 6A). The di-iron site of EncFtn<sub>SH</sub> is by necessity symmetrical, as it is formed through a dimer interface, while the FOC of ferritin does not have these constraints and varies in different species at a position equivalent to His65 of the second EncFtn monomer in the FOC interface (His65') (Figure 6A). Structural superimposition of the FOCs of ferritin and EncFtn brings the four-helix bundle of the ferritin fold into close alignment with the EncFtn dimer, showing that the two families of proteins have essentially the same architecture around the di-iron center (Figure 6B). The linker connecting helices 2 and 3 of ferritin is congruent with the start of the C-terminal helix of one EncFtn monomer and the N-terminal 3<sub>10</sub> helix of the second monomer (Figure 6C).

## Mass spectrometry of the EncFtn assembly

In order to confirm the assignment of the oligomeric state of EncFtn<sub>SH</sub> and investigate further the Fe<sup>2+</sup>-dependent assembly, we used native nano-electrospray ionization (nESI) and ion-mobility mass spectrometry (IM-MS). As described above, by recombinant production of EncFtn<sub>SH</sub> in minimal media we were able to limit the bioavailability of iron. Native MS analysis of EncFtn<sub>SH</sub> produced in



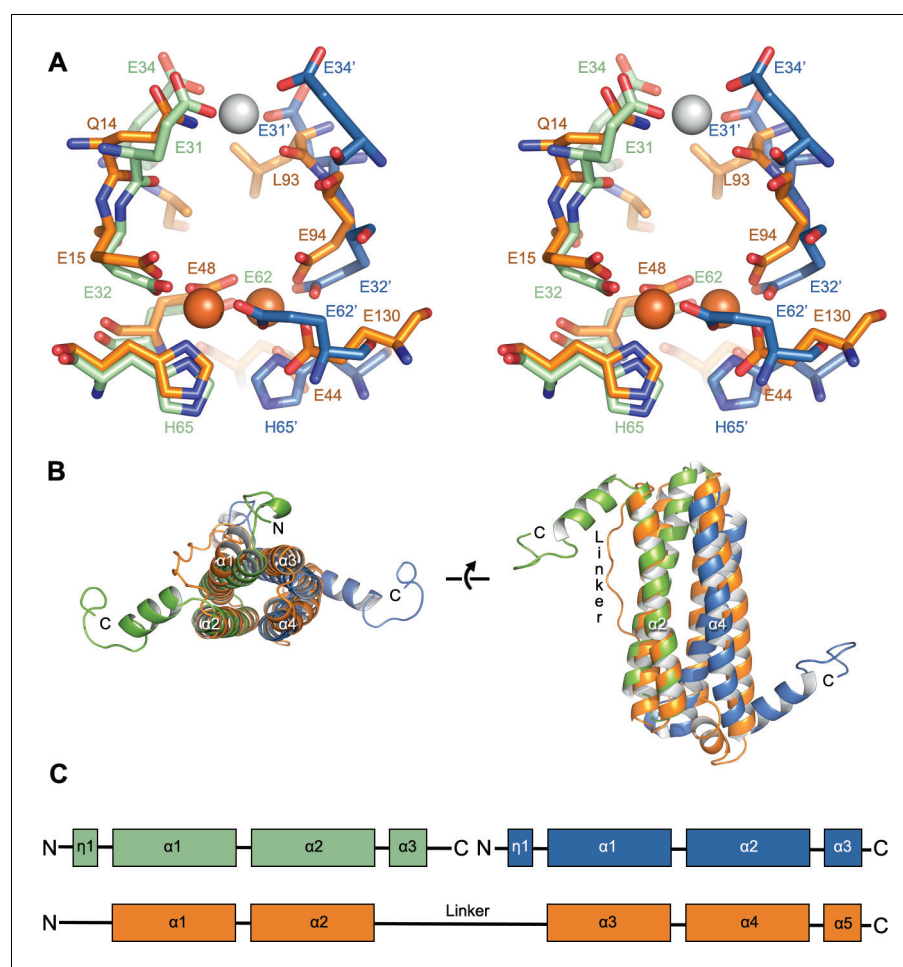
**Figure 5.** EncFtn<sub>SH</sub> metal binding sites. (A) Wall-eyed stereo view of the metal-binding dimerization interface of EncFtn<sub>SH</sub>. Protein residues are shown as sticks with blue and green carbons for the different subunits, iron ions are shown as orange spheres and calcium as grey spheres, and the glycolic acid ligand is shown with yellow carbon atoms coordinated above the di-iron center. The 2mFo-DFc electron density map is shown as a blue mesh contoured at 1.5  $\sigma$  and the NCS-averaged anomalous difference map is shown as an orange mesh and contoured at 10  $\sigma$ . (B) Iron coordination within the FOC including residues Glu32, Glu62, His65 and Tyr39 from two chains. Protein and metal ions are shown as in A. Coordination between the protein and iron ions is shown as yellow dashed lines with distances indicated. (C) Coordination of calcium within the dimer interface by four glutamic acid residues (E31 and E34 from two chains). The calcium ion is shown as a grey sphere and water molecules involved in the coordination of the calcium ion are shown as crosses. (D) Metal coordination site on the outer surface of EncFtn<sub>SH</sub>. The two calcium ions are coordinated by residues His57, Glu61 and Glu64 from the two chains of the FOC dimer, and are located at the outer surface of the complex, positioned 10 Å away from the FOC iron.

Figure 5 continued on next page

Figure 5 continued

DOI: [10.7554/eLife.18972.014](https://doi.org/10.7554/eLife.18972.014)

The following figure supplement is available for figure 5:

**Figure supplement 1.** Putative ligand-binding site in EncFtn<sub>sH</sub>.DOI: [10.7554/eLife.18972.015](https://doi.org/10.7554/eLife.18972.015)

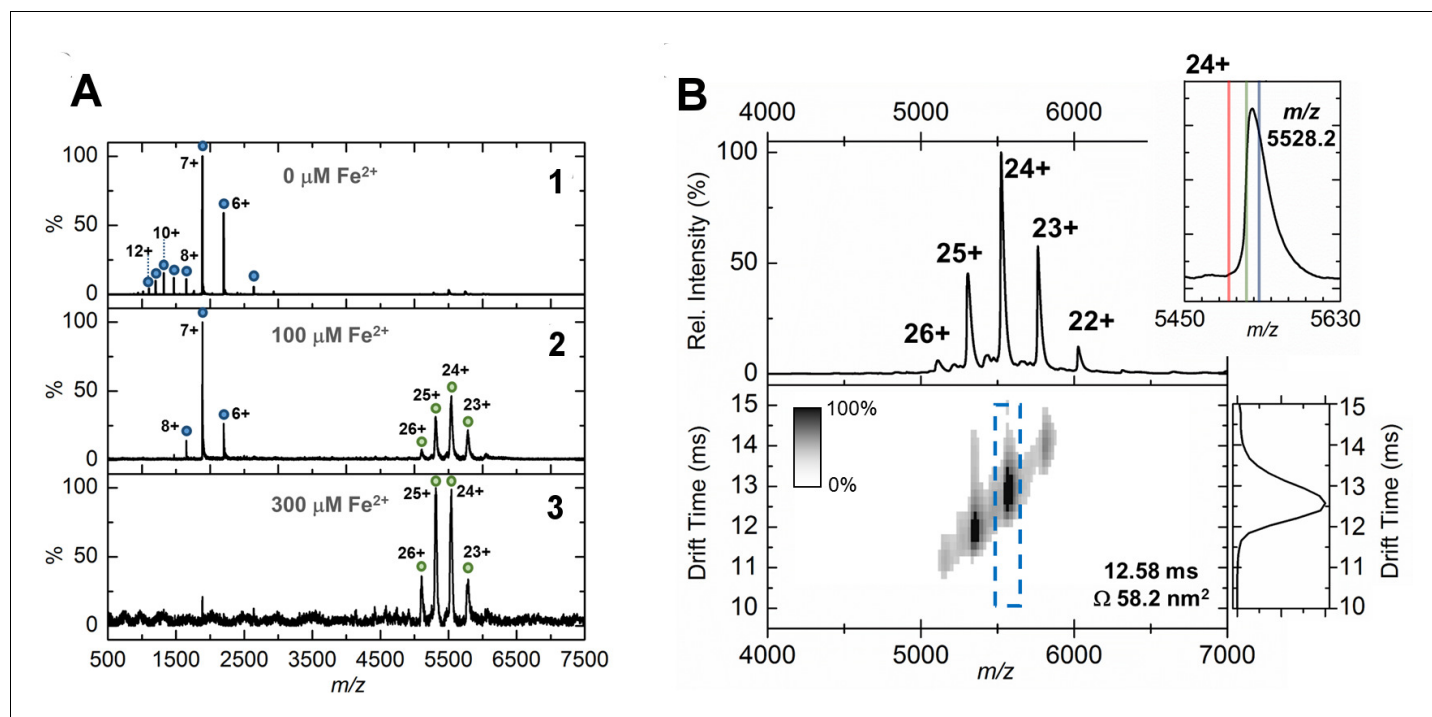
**Figure 6.** Comparison of the symmetric metal ion binding site of EncFtn<sub>sH</sub> and the ferritin FOC. (A) Structural alignment of the FOC residues in a dimer of EncFtn<sub>sH</sub> (green/blue) with a monomer of *Pseudo-nitzschia multiseriis* ferritin (PmFtn) (PDBID: 4ITW) (orange) (Pfaffen et al., 2013). Iron ions are shown as orange spheres and a single calcium ion as a grey sphere. Residues within the FOC are conserved between EncFtn and ferritin PmFtn, with the exception of residues in the position equivalent to H65' in the second subunit in the dimer (blue). The site in EncFtn with bound calcium is not present in other family members. (B) Secondary structure of aligned dimeric EncFtn<sub>sH</sub> and monomeric ferritin highlighting the conserved four-helix bundle. EncFtn<sub>sH</sub> monomers are shown in green and blue and aligned PmFtn monomer in orange as in A. (C) Cartoon of secondary structure elements in EncFtn dimer and ferritin. In the dimer of EncFtn that forms the FOC, the C-terminus of the first monomer (green) and N-terminus of the second monomer (blue) correspond to the position of the long linker between α2 and α3 in ferritin PmFtn.

DOI: [10.7554/eLife.18972.016](https://doi.org/10.7554/eLife.18972.016)

The following figure supplement is available for figure 6:

**Figure supplement 1.** Comparison of quaternary structure of EncFtn<sub>sH</sub> and ferritin.DOI: [10.7554/eLife.18972.017](https://doi.org/10.7554/eLife.18972.017)





**Figure 7.** Native mass spectrometry and ion mobility analysis of iron loading in EncFtn<sub>SH</sub>. All spectra were acquired in 100 mM ammonium acetate, pH 8.0 with a protein concentration of 5 μM. (A) Native nanoelectrospray ionization (nESI) mass spectrometry of EncFtn<sub>SH</sub> at varying iron concentrations. A1, nESI spectrum of iron-free EncFtn<sub>SH</sub> displays a charge state distribution consistent with EncFtn<sub>SH</sub> monomer (blue circles, 13,194 Da). Addition of 100 μM (A2) and 300 μM (A3) Fe<sup>2+</sup> results in the appearance of a second higher molecular weight charge state distribution consistent with a decameric assembly of EncFtn<sub>SH</sub> (green circles, 132.6 kDa). (B) Ion mobility (IM)-MS of the iron-bound holo-EncFtn<sub>SH</sub> decamer. Top, Peaks corresponding to the 22+ to 26+ charge states of a homo-decameric assembly of EncFtn<sub>SH</sub> are observed (132.6 kDa). Top Insert, Analysis of the 24+ charge state of the assembly at *m/z* 5528.2 Th. The theoretical average *m/z* of the 24+ charge state with no additional metals bound is marked by a red line (5498.7 Th); the observed *m/z* of the 24+ charge state indicates that the EncFtn<sub>SH</sub> assembly binds between 10 (green line, 5521.1 Th) and 15 Fe ions (blue line, 5532.4 Th) per decamer. Bottom, The arrival time distributions (ion mobility data) of all ions in the EncFtn<sub>SH</sub> charge state distribution displayed as a greyscale heat map (linear intensity scale). Bottom right, The arrival time distribution of the 24+ charge state (dashed blue box) has been extracted and plotted. The drift time for this ion is shown (ms), along with the calibrated collision cross section (CCS), Ω (nm<sup>2</sup>).

DOI: [10.7554/eLife.18972.018](https://doi.org/10.7554/eLife.18972.018)

The following figure supplements are available for figure 7:

**Figure supplement 1.** Native IM-MS analysis of the apo-EncFtn<sub>SH</sub> monomer.

DOI: [10.7554/eLife.18972.019](https://doi.org/10.7554/eLife.18972.019)

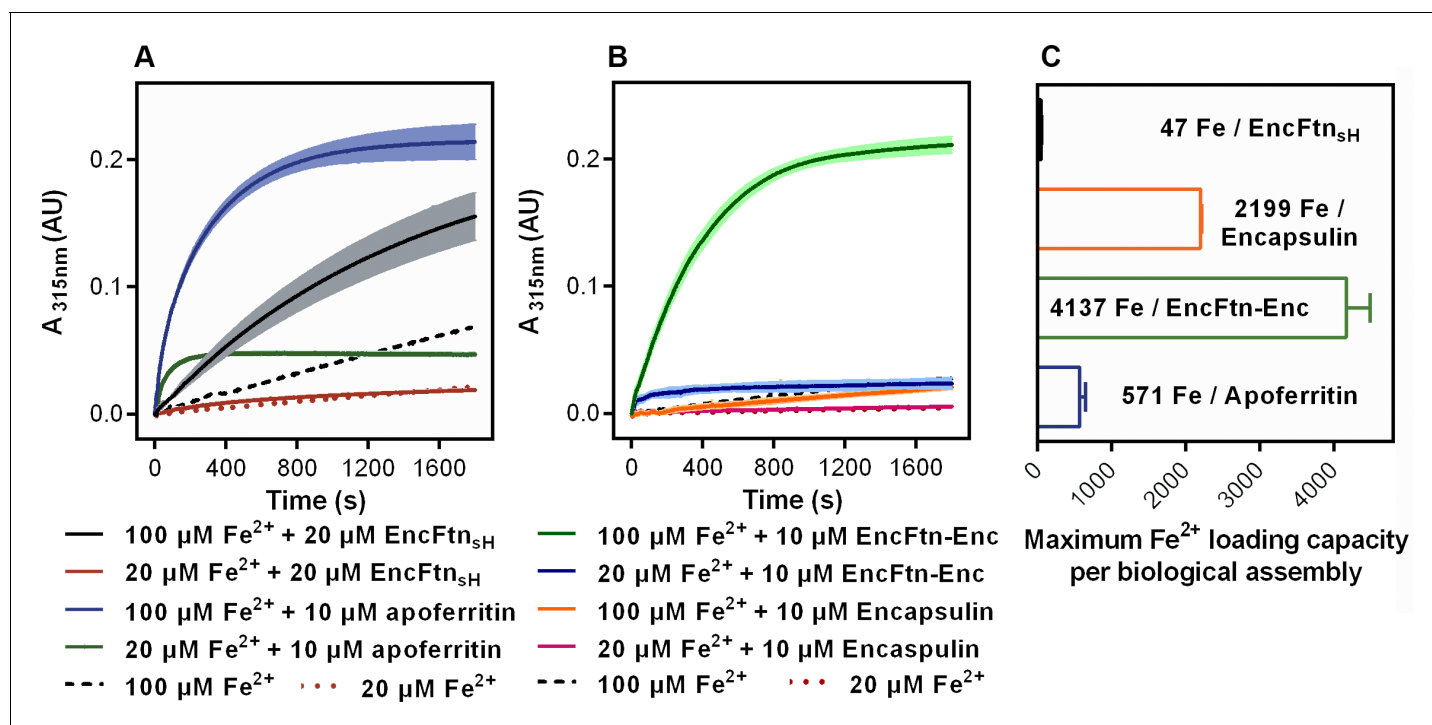
**Figure supplement 2.** Gas-phase disassembly of the holo-EncFtn<sub>SH</sub> decameric assembly.

DOI: [10.7554/eLife.18972.020](https://doi.org/10.7554/eLife.18972.020)

this way displayed a charge state distribution consistent with an EncFtn<sub>SH</sub> monomer (blue circles, **Figure 7A1**) with an average neutral mass of 13,194 Da, in agreement with the predicted mass of the EncFtn<sub>SH</sub> protein (13,194.53 Da). Under these conditions, no significant higher order assembly was observed and the protein did not have any coordinated metal ions. Titration with Fe<sup>2+</sup> directly before native MS analysis resulted in the appearance of a new charge state distribution, consistent with an EncFtn<sub>SH</sub> decameric assembly (+22 to +26; 132.65 kDa) (**Figure 7A2/3**). After instrument optimization, the mass resolving power achieved was sufficient to assign iron-loading in the complex to between 10 and 15 Fe ions per decamer (**Figure 7B**, inset top right), consistent with the presence of 10 irons in the FOC and the coordination of iron in the Glu31/34-site occupied by calcium in the crystal structure (Δmass observed ~0.67 kDa). MS analysis of EncFtn<sub>SH</sub> after addition of further Fe<sup>2+</sup> did not result in iron loading above this stoichiometry. Therefore, the extent of iron binding seen is limited to the FOC and Glu31/34 secondary metal binding site. These data suggest that the decameric assembly of EncFtn<sub>SH</sub> does not accrue iron in the same manner as classical ferritin, which is able to sequester around 4500 iron ions within its nanocage (Mann et al., 1986). Ion mobility

analysis of the EncFtn<sub>sH</sub> decameric assembly, collected with minimal collisional activation, suggested that it consists of a single conformation with a collision cross section (CCS) of 58.2 nm<sup>2</sup> (Figure 7B). This observation is in agreement with the calculated CCS of 58.7 nm<sup>2</sup> derived from our crystal structure of the EncFtn<sub>sH</sub> decamer (Marklund, 2015). By contrast, IM-MS measurements of the monomeric EncFtn<sub>sH</sub> at pH 8.0 under the same instrumental conditions revealed that the metal-free protein monomer exists in a wide range of charge states (+6 to +16) and adopts many conformations in the gas phase with collision cross sections ranging from 12 nm<sup>2</sup> to 26 nm<sup>2</sup> (Figure 7—figure supplement 1). These observations are indicative of an unstructured protein with little secondary or tertiary structure (Beveridge et al., 2014). Thus, IM-MS studies highlight that higher order structure in EncFtn<sub>sH</sub> is mediated/stabilized by metal binding, an observation that is in agreement with our solution studies. Taken together, these results suggest that di-iron binding, forming the FOC in EncFtn<sub>sH</sub>, is required to stabilize the 4-helix bundle dimer interface, essentially reconstructing the classical ferritin-like fold; once stabilized, these dimers readily associate as pentamers, and the overall assembly adopts the decameric ring arrangement observed in the crystal structure.

We subsequently performed gas phase disassembly of the decameric EncFtn<sub>sH</sub> using collision-induced dissociation (CID) tandem mass spectrometry. Under the correct CID conditions, protein assemblies can dissociate with retention of subunit and ligand interactions, and thus provide structurally-informative evidence as to the topology of the original assembly; this has been termed ‘atypical’ dissociation (Hall et al., 2013). For EncFtn<sub>sH</sub>, this atypical dissociation pathway was clearly



**Figure 8.** Spectroscopic evidence for the ferroxidase activity and comparison of iron loading capacity of apoferritin, EncFtn<sub>sH</sub>, encapsulin, and EncFtn-Enc. (A) Apoferritin (10 μM monomer concentration) and EncFtn<sub>sH</sub> decamer fractions (20 μM monomer concentration, 10 μM FOC concentration) were incubated with 20 and 100 μM iron (2 and 10 times molar equivalent Fe<sup>2+</sup> per FOC) and progress curves of the oxidation of Fe<sup>2+</sup> to Fe<sup>3+</sup> at 315 nm were recorded in a spectrophotometer. The background oxidation of iron at 20 and 100 μM in enzyme-free controls are shown for reference. (B) Encapsulin and EncFtn-Enc complexes at 10 μM asymmetric unit concentration were incubated with Fe<sup>2+</sup> at 20 and 100 μM and progress curves for iron oxidation at A<sub>315</sub> were measured in a UV/visible spectrophotometer. Enzyme free controls for background oxidation of Fe<sup>2+</sup> are shown for reference. (C) Histogram of the iron loading capacity per biological assembly of EncFtn<sub>sH</sub>, encapsulin, EncFtn-Enc and apoferritin. The results shown are for three technical replicates and represent the optimal iron loading by the complexes after three hours when incubated with Fe<sup>2+</sup>.

DOI: 10.7554/eLife.18972.021

The following figure supplement is available for figure 8:

**Figure supplement 1.** TEM visualization of iron-loaded bacterial nanocompartments and ferritin.

DOI: 10.7554/eLife.18972.022



evident; CID of the EncFtn<sub>sH</sub> decamer resulted in the appearance of a dimeric EncFtn<sub>sH</sub> subcomplex containing 0, 1, or 2 iron ions (**Figure 7—figure supplement 2**). In light of the crystal structure, this observation can be rationalized as dissociation of the EncFtn<sub>sH</sub> decamer by disruption of the non-FOC interface with at least partial retention of the FOC interface and the FOC-Fe. Thus, this observation supports our crystallographic assignment of the overall topology of the EncFtn<sub>sH</sub> assembly as a pentameric assembly of dimers with two iron ions located at the FOC dimer interface. In addition, this analysis provides evidence that the overall architecture of the complex is consistent in the crystal, solution and gas phases.

### Ferroxidase activity

In light of the identification of an iron-loaded FOC in the crystal structure of EncFtn and our native mass spectrometry data, we performed ferroxidase and peroxidase assays to demonstrate the catalytic activity of this protein. In addition, we also assayed equine apoferritin, an example of a classical ferritin enzyme, as a positive control. Unlike the Dps family of ferritin-like proteins, EncFtn showed no peroxidase activity when assayed with the substrate *ortho*-phenylenediamine (**Pesek et al., 2011**). The ferroxidase activity of EncFtn<sub>sH</sub> was measured by recording the progress curve of Fe<sup>2+</sup> oxidation to Fe<sup>3+</sup> at 315 nm after addition of 20 and 100 μM Fe<sup>2+</sup> (2 and 10 times molar ratio Fe<sup>2+</sup>/FOC). In both experiments the rate of oxidation was faster than background oxidation of Fe<sup>2+</sup> by molecular oxygen, and was highest for 100 μM Fe<sup>2+</sup> (**Figure 8A**). These data show that recombinant EncFtn<sub>sH</sub> acts as an active ferroxidase enzyme. When compared to apoferritin, EncFtn<sub>sH</sub> oxidized Fe<sup>2+</sup> at a slower rate and the reaction did not run to completion over the 1800 s of the experiment. Addition of higher quantities of iron resulted in the formation of a yellow/red precipitate at the end of the reaction. We also performed these assays on purified recombinant encapsulin; which, when assayed alone, did not display ferroxidase activity above background Fe<sup>2+</sup> oxidation (**Figure 8B**). In contrast, complexes of the full EncFtn encapsulin nanocompartment (i.e. the EncFtn-Enc protein complex) displayed ferroxidase activity comparable to apoferritin without the formation of precipitates (**Figure 8B**).

We attributed the precipitates observed in the EncFtn<sub>sH</sub> ferroxidase assay to the production of insoluble Fe<sup>3+</sup> complexes, which led us to propose that EncFtn does not directly store Fe<sup>3+</sup> in a mineral form. This observation agrees with native MS results, which indicates a maximum iron loading of 10–15 iron ions per decameric EncFtn; and the structure, which does not possess the enclosed iron-storage cavity characteristic of classical ferritins and Dps family proteins that can directly accrue mineralized Fe<sup>3+</sup> within their nanocompartment structures.

To analyze the products of these reactions and determine whether the EncFtn and encapsulin were able to store iron in a mineral form, we performed TEM on the reaction mixtures from the ferroxidase assay. The EncFtn<sub>sH</sub> reaction mixture showed the formation of large, irregular electron-dense precipitates (**Figure 8—figure supplement 1A**). A similar distribution of particles was observed after addition of Fe<sup>2+</sup> to the encapsulin protein (**Figure 8—figure supplement 1B**). In contrast, addition of Fe<sup>2+</sup> to the EncFtn-Enc nanocompartment resulted in small, highly regular, electron dense particles of approximately 5 nm in diameter (**Figure 8—figure supplement 1C**); we interpret these observations as controlled mineralization of iron within the nanocompartment. Addition of Fe<sup>2+</sup> to apoferritin resulted in a mixture of large particles and small (~2 nm) particles consistent with partial mineralization by the ferritin and some background oxidation of the iron (**Figure 8—figure supplement 1D**). Negative stain TEM of these samples revealed that upon addition of iron, the EncFtn<sub>sH</sub> protein showed significant aggregation (**Figure 8—figure supplement 1F**); while the encapsulin, EncFtn-Enc system, and apoferritin are present as distinct nanocompartments without significant protein aggregation (**Figure 8—figure supplement 1G–I**).

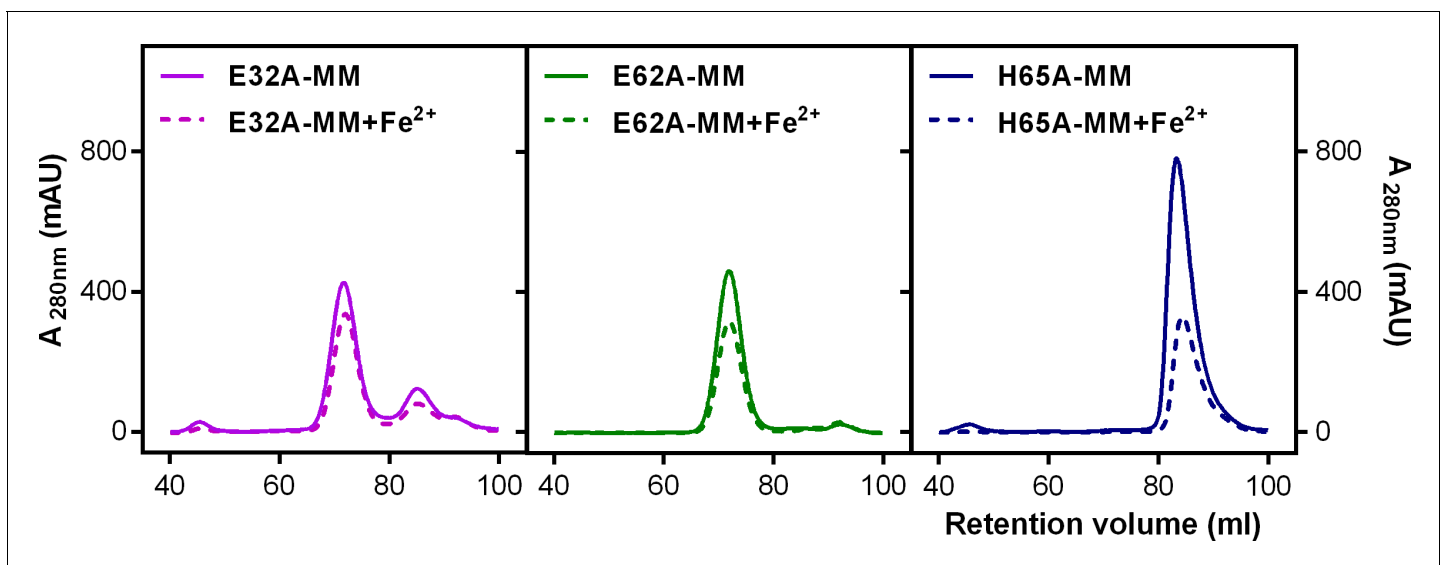
### Iron storage in encapsulin nanocompartments

The results of the ferroxidase assay and micrographs of the reaction products suggest that the oxidation and mineralization function of the classical ferritins are split between the EncFtn and encapsulin proteins, with the EncFtn acting as a ferroxidase and the encapsulin shell providing an environment and template for iron mineralization and storage. To investigate this further, we added Fe<sup>2+</sup> at various concentrations to samples of apo-ferritin, EncFtn, isolated encapsulin, and the EncFtn-Enc protein complex, and subjected these samples to a ferrozine assay to quantify the

amount of iron associated with the proteins after three hours of incubation. The maximum iron loading capacity of these systems was calculated as the quantity of iron per biological assembly (Figure 8C). In this assay, the EncFtn<sub>sH</sub> decamer binds a maximum of around 48 iron ions before excess iron induces protein precipitation. The encapsulin shell protein can sequester about 2200 iron ions before significant protein loss occurs, and the reconstituted EncFtn-Enc nanocompartment sequestered about 4150 iron ions. This latter result is significantly more than the apoferritin used in our assay, which sequesters approximately 570 iron ions in this assay (Figure 8C, Table 5).

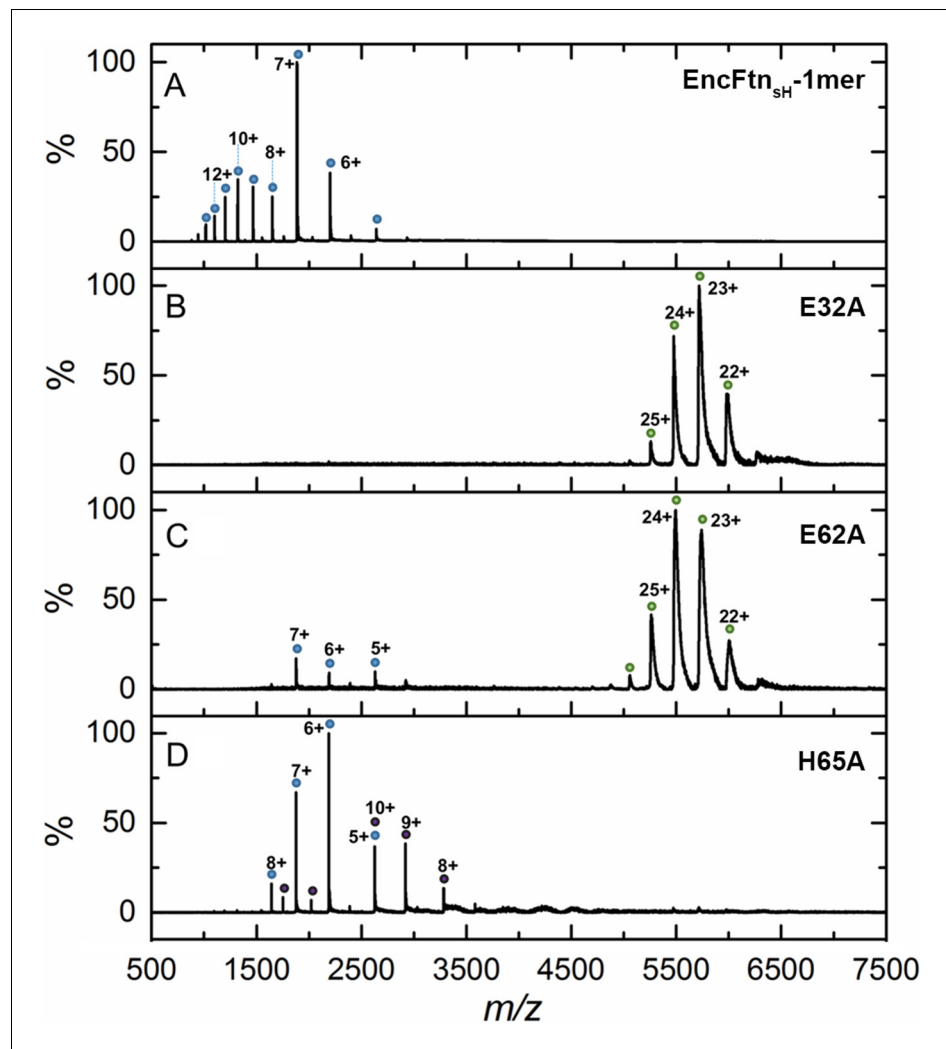
Consideration of the functional oligomeric states of these proteins, where EncFtn is a decamer and encapsulin forms an icosahedral cage, and estimation of the iron loading capacity of these complexes gives insight into the role of the two proteins in iron storage and mineralization. EncFtn decamers bind up to 48 iron ions (Figure 8C), which is significantly higher than the stoichiometry of fifteen metal ions visible in the FOC and E31/34-site of the crystal structure of the EncFtn<sub>sH</sub> decamer and our MS analysis. The discrepancy between these solution measurements and our MS analysis may indicate that there are additional metal-binding sites on the interior channel and exterior faces of the protein; this is consistent with our identification of a number of weak metal-binding sites at the surface of the protein in the crystal structure (Figure 5D). These observations are consistent with hydrated Fe<sup>2+</sup> ions being channeled to the active site from the E31/34-site and the subsequent exit of Fe<sup>3+</sup> products on the outer surface, as is seen in other ferritin family proteins (Pesek et al., 2011; Behera and Theil, 2014). While the isolated encapsulin shell does not display any ferroxidase activity, it binds around 2200 iron ions in our assay (Table 5). This implies that the shell can bind a significant amount of iron on its outer and inner surfaces. While the maximum reported loading capacity of classical ferritins is approximately 4500 iron ions (Mann et al., 1986), in our assay system we were only able to load apoferritin with around 570 iron ions. However, the recombinant EncFtn-Enc nanocompartment was able to bind over 4100 iron ions in the same time period, over seven times the amount seen for the apoferritin. We note we do not reach the experimental maximum iron loading for apoferritin and therefore the total iron-loading capacity of our system may be significantly higher than in this experimental system.

Taken together, our data show that EncFtn can catalytically oxidize Fe<sup>2+</sup> to Fe<sup>3+</sup>; however, iron binding in EncFtn is limited to the FOC and several surface metal binding sites. In contrast, the



**Figure 9.** Purification of recombinant *R. rubrum* EncFtn<sub>sH</sub> FOC mutants. Single mutants E32A, E62A, and H65A of EncFtn<sub>sH</sub> produced from *E. coli* BL21 (DE3) cells grown in MM and MM supplemented with iron were subjected to Superdex 200 size-exclusion chromatography. (A) Gel-filtration chromatogram of the E32A mutant form of EncFtn<sub>sH</sub> resulted in an elution profile with a majority of the protein eluting as the decameric form of the protein and a small proportion of monomer. (B) Gel-filtration chromatography of the E62A mutant form of EncFtn<sub>sH</sub> resulted in an elution profile with a single major decameric peak. (C) Gel-filtration chromatography of the H65A mutant form of EncFtn<sub>sH</sub> resulted in a single peak corresponding to the protein monomer.

DOI: 10.7554/eLife.18972.023



**Figure 10.** Native mass spectrometry of EncFtn<sub>SH</sub> mutants. All spectra were acquired in 100 mM ammonium acetate, pH 8.0 with a protein concentration of 5  $\mu$ M. (A) Wild-type EncFtn<sub>SH</sub> in the absence of iron displays a charge state distribution consistent with a monomer (see also [Figure 8](#)). (B) E32A EncFtn<sub>SH</sub> displays a charge states consistent with a decamer (green circles); a minor species, consistent with the monomer of E32A mutant is also observed (blue circles). (C) E62A EncFtn<sub>SH</sub> displays charge states consistent with a decamer (green circles). (D) H65A EncFtn<sub>SH</sub> displays charge states consistent with both monomer (blue circles) and dimer (purple circles). DOI: [10.7554/eLife.18972.024](https://doi.org/10.7554/eLife.18972.024)

encapsulin protein displays no catalytic activity, but has the ability to bind a considerable amount of iron. Finally, the EncFtn-Enc nanocompartment complex retains the catalytic activity of EncFtn, and sequesters iron within the encapsulin shell at a higher level than the isolated components of the system, and at a significantly higher level than the classical ferritins ([Andrews, 2010](#)). Furthermore, our recombinant nanocompartments may not have the physiological subunit stoichiometry, and the iron-loading capacity of native nanocompartments is potentially much higher than the level we have observed.

### Mutagenesis of the EncFtn<sub>SH</sub> ferroxidase center

To investigate the structural and biochemical role played by the metal binding residues in the di-iron FOC of EncFtn<sub>SH</sub> we produced alanine mutations in each of these residues: Glu32, Glu62, and His65. These EncFtn<sub>SH</sub> mutants were produced in *E. coli* cells grown in MM, both in the absence and presence of additional iron. The E32A and E62A mutants eluted from SEC at a volume consistent with the decameric form of EncFtn<sub>SH</sub>, with a small proportion of monomer; the H65A mutant eluted at a

**Table 4.** Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses. Friedel mates were averaged when calculating reflection numbers and statistics.

	WT	E32A	E62A	H65A
Data collection				
Wavelength (Å)	1.74	1.73	1.73	1.74
Resolution range (Å)	49.63 - 2.06 (2.10 - 2.06)	48.84 - 2.59 (2.683 - 2.59)	48.87 - 2.21 (2.29 - 2.21)	48.86 - 2.97 (3.08 - 2.97)
Space group	<i>P</i> 1 2 <sub>1</sub> 1	<i>P</i> 1 2 <sub>1</sub> 1	<i>P</i> 1 2 <sub>1</sub> 1	<i>P</i> 1 2 <sub>1</sub> 1
Unit cell (Å) a	98.18	97.78	98.09	98.03
b	120.53	120.28	120.23	120.29
c	140.30	140.53	140.36	140.43
β (°)	95.36	95.41	95.50	95.39
Total reflections	1,264,922 (41,360)	405,488 (36,186)	1,069,345 (95,716)	323,853 (32,120)
Unique reflections	197,873 (8,766)	100,067 (9,735)	162,379 (15,817)	66,658 (6,553)
Multiplicity	6.4 (4.7)	4.1 (3.7)	6.6 (6.1)	4.9 (4.9)
Anomalous multiplicity	3.2 (2.6)	N/A	N/A	N/A
Completeness (%)	99.2 (88.6)	99.0 (97.0)	100 (97.0)	100 (99.0)
Anomalous completeness (%)	96.7 (77.2)	N/A	N/A	N/A
Mean I/sigma(I)	10.6 (1.60)	8.46 (1.79)	13.74 (1.80)	8.09 (1.74)
Wilson B-factor	26.98	40.10	33.97	52.20
R <sub>merge</sub>	0.123 (0.790)	0.171 (0.792)	0.0979 (1.009)	0.177 (0.863)
R <sub>meas</sub>	0.147 (0.973)	0.196 (0.923)	0.1064 (1.107)	0.199 (0.966)
CC1/2	0.995 (0.469)	0.985 (0.557)	0.998 (0.642)	0.989 (0.627)
CC*	0.999 (0.846)	0.996 (0.846)	0.999 (0.884)	0.997 (0.878)
Image DOI	10.7488/ds/1342	10.7488/ds/1419	10.7488/ds/1420	10.7488/ds/1421
Refinement				
R <sub>work</sub>	0.171 (0.318)	0.183 (0.288)	0.165 (0.299)	0.186 (0.273)
R <sub>free</sub>	0.206 (0.345)	0.225 (0.351)	0.216 (0.364)	0.237 (0.325)
Number of non-hydrogen atoms	23,222	22,366	22,691	22,145
macromolecules	22,276	22,019	21,965	22,066
ligands	138	8	24	74
water	808	339	702	5
Protein residues	2,703	2,686	2,675	2,700
RMS(bonds) (Å)	0.012	0.005	0.011	0.002
RMS(angles) (°)	1.26	0.58	1.02	0.40
Ramachandran favored (%)	100	99	100	99
Ramachandran allowed (%)	0	1	0	1
Ramachandran outliers (%)	0	0	0	0
Clash score	1.42	1.42	1.79	0.97
Average B-factor (Å <sup>2</sup> )	33.90	42.31	41.34	47.68
macromolecules	33.80	42.35	41.31	47.60
ligands	40.40	72.80	65.55	72.34
solvent	36.20	38.95	41.46	33.85
PDB ID	5DA5	5L89	5L8B	5L8G

DOI: [10.7554/eLife.18972.025](https://doi.org/10.7554/eLife.18972.025)

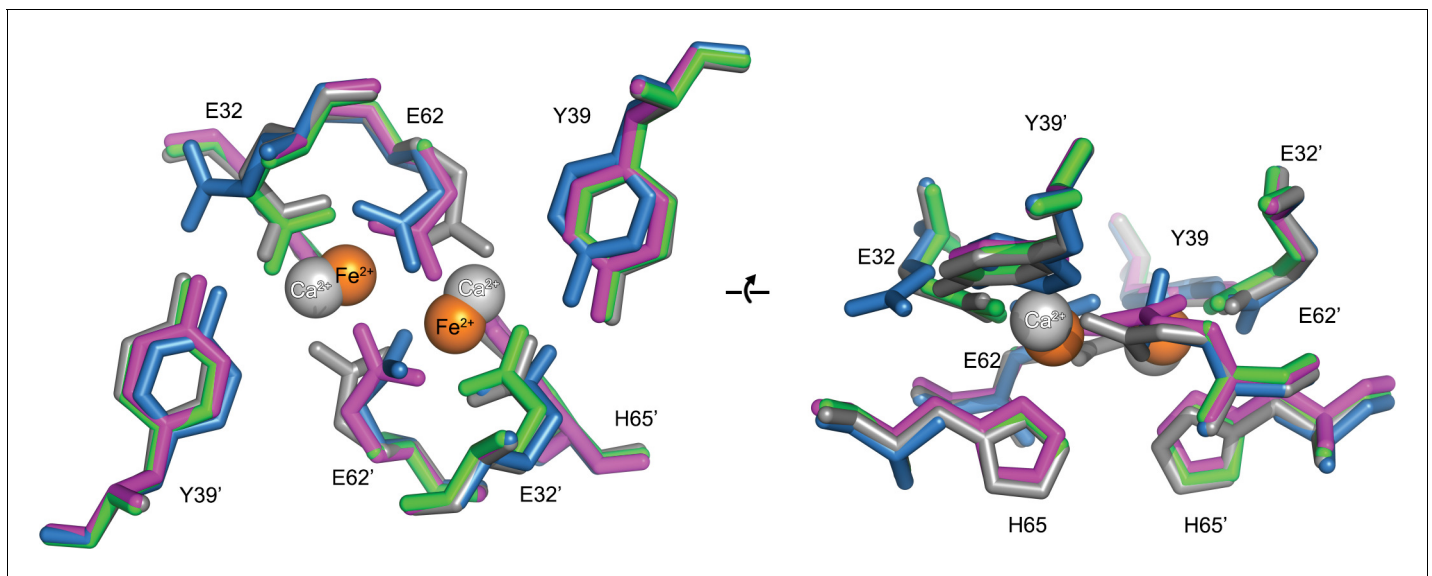
**Table 5.** Iron loading capacity of EncFtn, encapsulin and ferritin. Protein samples (at 8.5  $\mu$ M) including decameric EncFtn<sub>sH</sub>, encapsulin, EncFtn-Enc and apoferritin were mixed with Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>) (in 0.1% (v/v) HCl) of different concentrations in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl buffer at room temperature for 3 hrs in the air. Protein-Fe mixtures were centrifuged at 13,000 x g to remove precipitated material and desalted prior to the Fe and protein content analysis by ferrozine assay and BCA microplate assay, respectively. Fe to protein ratio was calculated to indicate the Fe binding capacity of the protein. Protein stability was compromised at high iron concentrations; therefore, the highest iron loading with the least protein precipitation was used to derive the maximum iron loading capacity per biological assembly (underlined and highlighted in bold). The biological unit assemblies are a decamer for EncFtn<sub>sH</sub>, a 60mer for encapsulin, a 60mer of encapsulin loaded with 12 copies of decameric EncFtn in the complex, and 24mer for horse spleen apoferritin. Errors are quoted as the standard deviation of three technical repeats in both the ferrozine and BCA microplate assays. The proteins used in Fe loading experiment came from a single preparation.

Protein sample	Fe(NH <sub>4</sub> ) <sub>2</sub> (SO <sub>4</sub> ) loading ( $\mu$ M)	Fe detected by ferrozine assay ( $\mu$ M)	Protein detected by BCA microplate assay ( $\mu$ M)	Fe / monomeric protein	Maximum Fe loading per biological assembly unit
8.46 $\mu$ M EncFtn <sub>sH</sub> -10mer	0	4.73 $\pm$ 2.32	5.26 $\pm$ 0.64	0.90 $\pm$ 0.44	
	39.9	9.93 $\pm$ 1.20	5.36 $\pm$ 0.69	1.85 $\pm$ 0.22	
	84	17.99 $\pm$ 2.01	4.96 $\pm$ 0.04	3.63 $\pm$ 0.41	
	147	21.09 $\pm$ 1.94	4.44 $\pm$ 0.21	4.75 $\pm$ 0.44	48 $\pm$ 4
	224	28.68 $\pm$ 0.30	3.73 $\pm$ 0.53	7.68 $\pm$ 0.08	
	301	11.27 $\pm$ 1.10	2.50 $\pm$ 0.05	4.51 $\pm$ 0.44	
8.50 $\mu$ M Encapsulin	0	-1.02 $\pm$ 0.54	8.63 $\pm$ 0.17	-0.12 $\pm$ 0.06	
	224	62.24 $\pm$ 2.49	10.01 $\pm$ 0.58	6.22 $\pm$ 0.35	
	301	67.94 $\pm$ 3.15	8.69 $\pm$ 0.42	7.81 $\pm$ 0.36	
	450	107.96 $\pm$ 8.88	8.50 $\pm$ 0.69	12.71 $\pm$ 1.05	
	700	97.51 $\pm$ 3.19	7.26 $\pm$ 0.20	13.44 $\pm$ 0.44	
	1000	308.63 $\pm$ 2.06	8.42 $\pm$ 0.34	36.66 $\pm$ 0.24	2199 $\pm$ 15
	1500	57.09 $\pm$ 0.90	1.44 $\pm$ 0.21	39.77 $\pm$ 0.62	
	2000	9.2 $\pm$ 1.16	0.21 $\pm$ 0.14	44.73 $\pm$ 5.63	
8.70 $\mu$ M EncFtn- Enc	0	3.31 $\pm$ 1.57	6.85 $\pm$ 0.07	0.48 $\pm$ 0.23	
	224	116.27 $\pm$ 3.74	7.63 $\pm$ 0.12	15.25 $\pm$ 0.49	
	301	132.86 $\pm$ 4.03	6.66 $\pm$ 0.31	19.96 $\pm$ 0.61	
	450	220.57 $\pm$ 27.33	6.12 $\pm$ 1.07	36.06 $\pm$ 4.47	
	700	344.03 $\pm$ 40.38	6.94 $\pm$ 0.17	49.58 $\pm$ 5.82	
	1000	496.00 $\pm$ 38.48	7.19 $\pm$ 0.08	68.94 $\pm$ 5.35	4137 $\pm$ 321
	1500	569.98 $\pm$ 73.63	5.73 $\pm$ 0.03	99.44 $\pm$ 12.84	
	2000	584.30 $\pm$ 28.33	4.88 $\pm$ 0.22	119.62 $\pm$ 5.80	
8.50 $\mu$ M Apoferritin	0	3.95 $\pm$ 2.26	9.37 $\pm$ 0.24	0.42 $\pm$ 0.25	
	42.5	10.27 $\pm$ 1.12	8.27 $\pm$ 0.30	1.24 $\pm$ 0.18	
	212.5	44.48 $\pm$ 2.76	7.85 $\pm$ 0.77	5.67 $\pm$ 0.83	
	637.5	160.93 $\pm$ 4.27	6.76 $\pm$ 0.81	23.79 $\pm$ 3.12	571 $\pm$ 75
	1275	114.92 $\pm$ 3.17	3.84 $\pm$ 0.30	29.91 $\pm$ 2.95	
	1700	91.40 $\pm$ 3.37	3.14 $\pm$ 0.35	29.13 $\pm$ 3.86	

DOI: 10.7554/eLife.18972.026

volume consistent with the monomeric form of EncFtn<sub>sH</sub> (**Figure 9**). For all of the mutants studied, no change in oligomerization state was apparent upon addition of Fe<sup>2+</sup> *in vitro*.

In addition to SEC studies, native mass spectrometry of the apo-EncFtn<sub>sH</sub> mutants was performed and compared with the wild-type apo-EncFtn<sub>sH</sub> protein (**Figure 10**). As described above, the apo-EncFtn<sub>sH</sub> has a charge state distribution consistent with an unstructured monomer, and decamer formation is only initiated upon addition of ferrous iron. Both the E32A mutant and E62A mutant displayed charge state distributions consistent with decamers, even in the absence of Fe<sup>2+</sup>. This gas-



**Figure 11.** Comparison of the EncFtn<sub>SH</sub> FOC mutants vs wild type. The structures of the three EncFtn<sub>SH</sub> mutants were all determined by X-ray crystallography. The E32A, E62A and H65A mutants were crystallized in identical conditions to the wild type. EncFtn<sub>SH</sub> structure and were essentially isomorphous in terms of their unit cell dimensions. The FOC residues of the mutants and native EncFtn<sub>SH</sub> structures are shown as sticks with coordinated Fe<sup>2+</sup> as orange and Ca<sup>2+</sup> as grey spheres and are colored as follows: wild type, grey; E32A, pink; E62A, green; H65A, blue. Of the mutants, only H65A has any coordinated metal ions, which appear to be calcium ions from the crystallization condition. The overall organization of FOC residues is retained in the mutants, with almost no backbone movements. Significant differences center around Tyr39, which moves to coordinate the bound calcium ions in the H65A mutant; and Glu32, which moves away from the metal ions in this structure.

DOI: [10.7554/eLife.18972.027](https://doi.org/10.7554/eLife.18972.027)

The following figure supplements are available for figure 11:

**Figure supplement 1.** FOC dimer interface of EncFtn<sub>SH</sub>-E32A mutant.

DOI: [10.7554/eLife.18972.028](https://doi.org/10.7554/eLife.18972.028)

**Figure supplement 2.** FOC dimer interface of EncFtn<sub>SH</sub>-E62A mutant.

DOI: [10.7554/eLife.18972.029](https://doi.org/10.7554/eLife.18972.029)

**Figure supplement 3.** FOC dimer interface of EncFtn<sub>SH</sub>-H65A mutant.

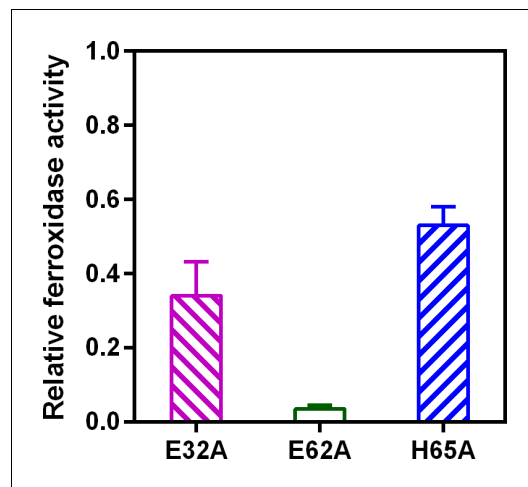
DOI: [10.7554/eLife.18972.030](https://doi.org/10.7554/eLife.18972.030)

phase observation is consistent with SEC measurements, which indicate both of these variants were also decamers in solution. Thus it seems that these mutations allow the decamer to form in the absence of iron in the FOC. In contrast to the glutamic acid mutants, MS analysis of the H65A mutant is similar to wild-type apo-EncFtn<sub>SH</sub> and is present as a monomer; interestingly a minor population of dimeric H65A was also observed.

We propose that the observed differences in the oligomerization state of the E32A and E62A mutants compared to wild-type are due to the changes in the electrostatic environment within the FOC. At neutral pH the glutamic acid residues are negatively charged, while the histidine residues are predominantly in their uncharged state. In the wild-type (WT) EncFtn<sub>SH</sub> this leads to electrostatic repulsion between subunits in the absence of iron. Coordination of Fe<sup>2+</sup> in this site stabilizes the dimer and reconstitutes the active FOC. The geometric arrangement of Glu32 and Glu62 in the FOC explains their behavior in solution and the gas phase, where they both favor the formation of decamers due to the loss of a repulsive negative charge. The FOC in the H65A mutant is destabilized through the loss of this metal coordinating residue and potential positive charge carrier, thus favoring the monomer in solution and the gas phase.

To understand the impact of the mutants on the organization and metal binding of the FOC, we determined the X-ray crystal structures of each of the EncFtn<sub>SH</sub> mutants (See **Table 4** for data collection and refinement statistics). The crystal packing of all of the mutants in this study is essentially isomorphous to the EncFtn<sub>SH</sub> structure. All of the mutants display the same decameric arrangement in the crystals as the EncFtn<sub>SH</sub> structure, and the monomers superimpose with an average RMSD<sub>Cα</sub> of less than 0.2 Å.





**Figure 12.** Relative ferroxidase activity of EncFtn<sub>SH</sub> mutants. EncFtn<sub>SH</sub> and the mutant forms E32A, E62A and H65A, each at 20  $\mu$ M, were mixed with 100  $\mu$ M acidic Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>. Ferroxidase activity of the mutant forms is determined by measuring the absorbance at 315 nm for 1800 s at 25  $^{\circ}$ C as an indication of Fe<sup>3+</sup> formation. The relative ferroxidase activity of mutants is plotted as a proportion of the activity of the wild-type protein using the endpoint measurement of A<sub>315</sub>. Three technical repeats were performed and the plotted error bars represent the calculated standard deviations. The FOC mutants showed reduced ferroxidase activity to varied extents, among which E62A significantly abrogated the ferroxidase activity.

DOI: [10.7554/eLife.18972.031](https://doi.org/10.7554/eLife.18972.031)

The following figure supplement is available for figure 12:

**Figure supplement 1.** Progress curves recording ferroxidase activity of EncFtn<sub>SH</sub> mutants. 20  $\mu$ M wild-type EncFtn<sub>SH</sub>, E32A, E62A and H65A mutants were mixed with 20  $\mu$ M or 100  $\mu$ M acidic Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>, respectively.

DOI: [10.7554/eLife.18972.032](https://doi.org/10.7554/eLife.18972.032)

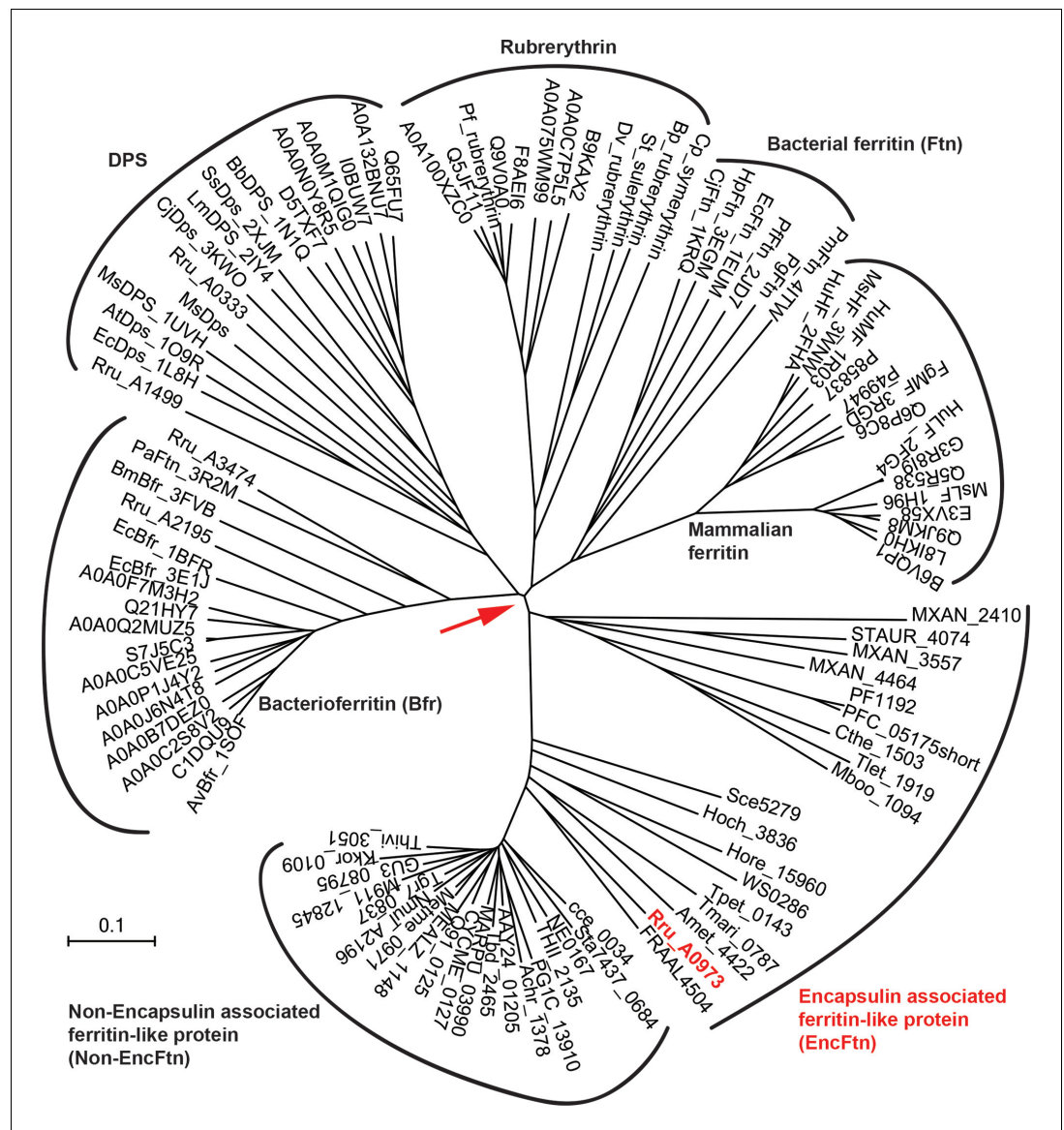
Close inspection of the region of the protein around the FOC in each of the mutants highlights their effect on metal binding (**Figure 11** and **Figure 11—figure supplement 1–3**). In the E32A mutant the position of the side chains of the remaining iron coordinating residues in the FOC is essentially unchanged, but the absence of the axial-metal coordinating ligand provided by the Glu32 side chain abrogates metal binding in this site. The Glu31/34-site also lacks metal, with the side chain of Glu31 rotated by 180 $^{\circ}$  at the C $_{\beta}$  in the absence of metal (**Figure 11—figure supplement 1**). The E62A mutant has a similar effect on the FOC to the E32A mutant, however the entry site still has a calcium ion coordinated between residues Glu31 and Glu34 (**Figure 11—figure supplement 2**). The H65A mutant diverges significantly from the wild type in the position of the residues Glu32 and Tyr39 in the FOC. E32 appears in either the original orientation as the wild type and coordinates Ca<sup>2+</sup> in this position, or it is flipped by 180 $^{\circ}$  at the C $_{\beta}$ , moving away from the coordinated calcium ion in the FOC. Tyr39 moves closer to Ca<sup>2+</sup> compared to the wild-type and coordinates the calcium ion (**Figure 11—figure supplement 3**). A single calcium ion is present in the entry site of this mutant; however, Glu31 of one chain is rotated away from the metal ion and is not involved in coordination.

Taken together the results of our data show that these changes to the FOC of EncFtn still permit the formation of the decameric form of the protein. While the proteins all appear decameric in crystals, their solution and gas-phase behavior differs considerably and the mutants no longer show metal-dependent oligomerization. These results highlight the importance of metal coordination in the FOC for the stability and assembly of the EncFtn protein.

To address the question of how mutagenesis of the iron coordinating residues affects the enzymatic activity of the EncFtn<sub>SH</sub> protein we recorded progress curves for the oxidation of Fe<sup>2+</sup> to Fe<sup>3+</sup> by the different mutants as before. Mutagenesis of E32A and H65A reduces the activity of EncFtn<sub>SH</sub> by about 40%–55%; the E62A mutant completely abrogates activity, presumably through the loss of the bridging coordination for the formation of the di-nuclear iron center of the FOC (**Figure 12**). Collectively, the effect of mutating these residues in the FOC confirms the importance of the iron coordinating residues for the ferroxidase activity of the EncFtn<sub>SH</sub> protein.

## Discussion

Our study reports on a new class of ferritin-like proteins (EncFtn), which are associated with bacterial encapsulin nanocompartments (Enc). By studying the EncFtn from *R. rubrum* we demonstrate that iron binding results in assembly of EncFtn decamers, which display a unique annular architecture. Despite a radically different quaternary structure to the classical ferritins, the four-helical bundle scaffold and FOC of EncFtn<sub>SH</sub> are strikingly similar to ferritin (**Figure 6A**). A sequence-based



**Figure 13.** Phylogenetic tree of ferritin family proteins. The tree was built using the Neighbor-Joining method (Saitou and Nei, 1987) based on step-wise amino acid sequence alignment of the four-helical bundle portions of ferritin family proteins (Supplementary file 1). The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree; the likely root of the tree is indicated by a red arrow. The evolutionary distances were computed using the p-distance method (Nei and Kumar, 2000) and are in the units of the number of amino acid differences per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 2.5). The analysis involved 104 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 262 positions in the final dataset. Evolutionary analyses were conducted in MEGA7 (McCoy et al., 2007)

DOI: 10.7554/eLife.18972.033

phylogenetic tree for proteins in the ferritin family was constructed; in addition to the classical ferritins, bacterioferritins and Dps proteins, our analysis included the encapsulin-associated ferritin-like proteins (EncFtns) and a group related to these, but lacking the encapsulin sequence (Non-EncFtn). The analysis revealed that the EncFtn and Non-EncFtn proteins form groups distinct from the other clearly delineated groups of ferritins, and represent outliers in the tree (Figure 13). While it is difficult to infer ancestral lineages in protein families, the similarity seen in the active site scaffold of these proteins highlights a shared evolutionary relationship between EncFtn proteins and other

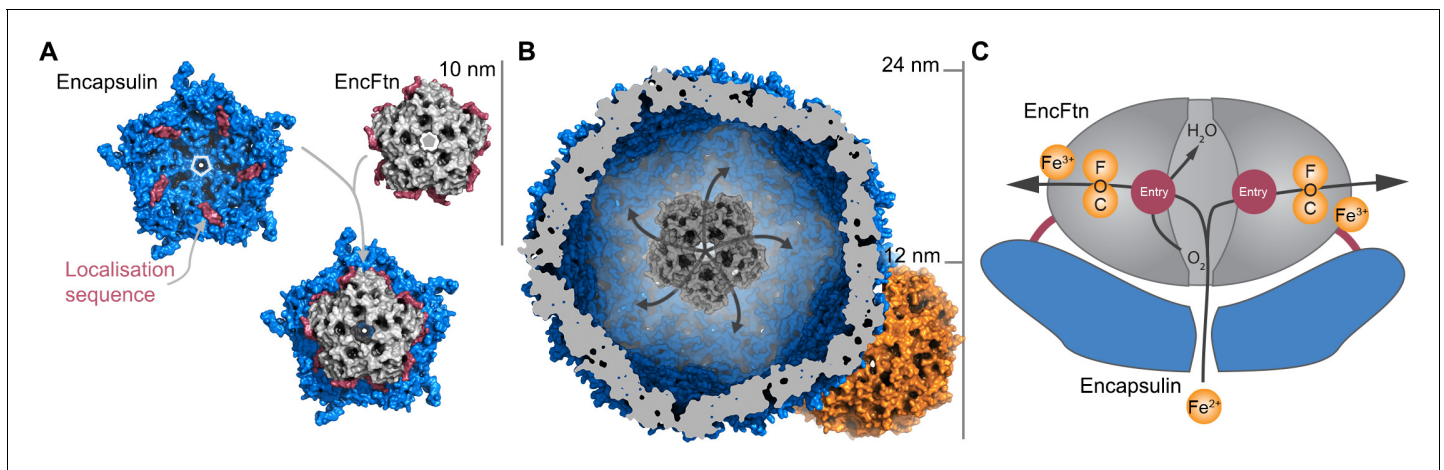
members of the ferritin superfamily that has been noted in previous studies (Andrews, 2010; Lundin et al., 2012). From this analysis, we propose that the four-helical fold of the classical ferritins may have arisen through gene duplication of an ancestor of EncFtn. This gene duplication would result in the C-terminal region of one EncFtn monomer being linked to the N-terminus of another and thus stabilizing the four-helix bundle fold within a single polypeptide chain (Figure 6B). Linking the protein together in this way relaxes the requirement for the maintenance of a symmetrical FOC and thus provides a path to the diversity in active-site residues seen across the ferritin family (Figure 6A, residues Glu95, Gln128 and Glu131 in PmFtn, Supplementary file 1) (Andrews, 2010; Lundin et al., 2012).

## Relationship between ferritin structure and activity

The quaternary arrangement of classical ferritins into an octahedral nanocage and Dps into a dodecamer is absolutely required for their function as iron storage compartments (Chasteen and Harrison, 1999). The oxidation and mineralization of iron must be spatially separated from the host cytosol to prevent the formation of damaging hydroxyl radicals in the Fenton and Haber-Weiss reactions (Honarmand Ebrahimi et al., 2012). This is achieved in all ferritins by confining the oxidation of iron to the interior of the protein complex, thus achieving sequestration of the  $\text{Fe}^{3+}$  mineralization product. A structural alignment of the FOC of EncFtn with the classical ferritin PmFtn shows that the central ring of EncFtn corresponds to the external surface of ferritin, while the outer circumference of EncFtn is congruent with the inner mineralization surface of ferritin (Figure 6—figure supplement 1A). This overlay highlights the fact that the ferroxidase center of EncFtn faces in the opposite direction relative to the classical ferritins and is essentially inside out regarding iron storage space (Figure 6—figure supplement 1B, boxed region). Analysis of each of the single mutations (E32A, E62A and H65A) made in the FOC highlights the importance of the iron-coordinating residues in the catalytic activity of EncFtn. Furthermore, the position of the calcium ion coordinated by Glu31 and Glu34 seen in the EncFtn<sub>sH</sub> structure suggests an entry site to channel metal ions into the FOC; we propose that this site binds hydrated iron ions *in vivo* and acts as a selectivity filter and gate for the FOC (Halder et al., 2011). The constellation of charged residues on the outer circumference of EncFtn (His57, Glu61 and Glu64) could function in the same way as the residues lining the mineralization surface within the classical ferritin nanocage (Le Brun et al., 2010), and given their proximity to the FOC these sites may be the exit portal and mineralization site (Honarmand Ebrahimi et al., 2012).

The absolute requirement for the spatial separation of oxidation and mineralization in ferritins suggests that the EncFtn family proteins are not capable of storing iron minerals due to the absence of an enclosed compartment in their structure (Figure 6—figure supplement 1B). Our biochemical characterization of EncFtn supports this hypothesis, indicating that while this protein is capable of oxidizing iron, it does not accrue mineralized iron in an analogous manner to classical ferritins. While EncFtn does not store iron itself, its association with the encapsulin nanocage suggests that mineralization occurs within the cavity of the encapsulin shell (McHugh et al., 2014). Our ferroxidase assay data on the recombinant EncFtn-Enc nanocompartments, which accrue over 4100 iron ions per complex and form regular nanoparticles, are consistent with the encapsulin protein acting as the store for iron oxidized by the EncFtn enzyme. TEM analysis of the reaction products shows the production of homogeneous iron nanoparticles only in the EncFtn-Enc nanocompartment (Figure 8—figure supplement 1).

Docking the decamer structure of EncFtn<sub>sH</sub> into the pentamer of the *T. maritima* encapsulin Tmari\_0786 (PDB ID: 3DKT) (Sutter et al., 2008) shows that the position of the C-terminal extensions of our EncFtn<sub>sH</sub> structure are consistent with the localization sequences seen bound to the encapsulin protein (Figure 14A). Thus, it appears that the EncFtn decamer is the physiological state of this protein. This arrangement positions the central ring of EncFtn directly above the pore at the five-fold symmetry axis of the encapsulin shell and highlights a potential route for the entry of iron into the encapsulin and towards the active site of EncFtn. A comparison of the encapsulin nanocompartment and the ferritin nanocage highlights the size differential between the two complexes (Figure 14B) that allows the encapsulin to store significantly more iron. The presence of five FOCs per EncFtn<sub>sH</sub> decamer and the fact that the icosahedral encapsulin nanocage can hold up to twelve of decameric EncFtn between each of the internal five-fold vertices means that they can achieve a high rate of iron mineralization across the entire nanocompartment. This arrangement of multiple



**Figure 14.** Model of iron oxidation in encapsulin nanocompartments. (A) Model of EncFtn<sub>SH</sub> docking to the encapsulin shell. A single pentamer of the icosahedral *T. maritima* encapsulin structure (PDBID: 3DKT) (Sutter et al., 2008) is shown as a blue surface with the encapsulin localization sequence of EncFtn shown as a purple surface. The C-terminal regions of the EncFtn subunits correspond to the position of the localization sequences seen in 3DKT. Alignment of EncFtn<sub>SH</sub> with 3DKT positions the central channel directly above the pore in the 3DKT pentamer axis (shown as a grey pentagon). (B) Surface view of EncFtn within the encapsulin nanocompartment (grey and blue respectively). The lumen of the encapsulin nanocompartment is considerably larger than the interior of ferritin (shown in orange behind the encapsulin for reference) and thus allows the storage of significantly more iron. The proposed pathway for iron movement through the encapsulin shell and EncFtn FOC is shown with arrows. (C) Model of iron oxidation within an encapsulin nanocompartment. As EncFtn is unable to mineralize iron on its surface directly, Fe<sup>2+</sup> must pass through the encapsulin shell to access the first metal binding site within the central channel of EncFtn<sub>SH</sub> (entry site) prior to oxidation within the FOC and release as Fe<sup>3+</sup> to the outer surface of the protein where it can be mineralized within the lumen of the encapsulin cage.

DOI: 10.7554/eLife.18972.034

reaction centers in a single protein assembly is reminiscent of classical ferritins, which has 24 FOCs distributed around the nanocage.

Our structural data, coupled with biochemical and ICP-MS analysis, suggest a model for the activity of the encapsulin iron-megastore (Figure 14C). The crystal structure of the *T. maritima* encapsulin shell protein has a negatively charged pore positioned to allow the passage of Fe<sup>2+</sup> into the encapsulin and directs the metal towards the central, negatively charged hole of the EncFtn ring (Figure 4—figure supplement 1). The five metal-binding sites on the interior of the ring (Glu31/34-sites) may select for the Fe<sup>2+</sup> ion and direct it towards their cognate FOCs. We propose that the oxidation of Fe<sup>2+</sup> to Fe<sup>3+</sup> occurs within the FOC according to the model postulated by (Honarmand Ebrahimi et al., 2012) in which the FOC acts as a substrate site through which iron passes and is released on to weakly coordinating sites at the outer circumference of the protein (His57, Glu61 and Glu64), where it is able to form ferrihydrite minerals which can be safely deposited within the lumen of the encapsulin nanocompartment (Figure 14).

Here we describe for the first time the structure and biochemistry of a new class of encapsulin-associated ferritin-like protein and demonstrate that it has an absolute requirement for compartmentalization within an encapsulin nanocage to act as an iron store. Further work on the EncFtn-Enc nanocompartment will establish the structural basis for the movement of iron through the encapsulin shell, the mechanism of iron oxidation by the EncFtn FOC and its subsequent storage in the lumen of the encapsulin nanocompartment.

## Materials and methods

### Cloning

Genes of interest were amplified by PCR using *R. rubrum* ATCC 11,170 genomic DNA (DSMZ) as the template and KOD Hot Start DNA Polymerase (Novagen). Primers used in this study are listed in **Supplementary file 2**. PCR products were visualized in 0.8% agarose gel stained with SYBR Safe (Life Technologies, UK). Fragments of interest were purified by gel extraction (Qiagen, UK) before digestion by endonuclease restriction enzymes (Thermo Fisher Scientific, UK) at 37°C for 1 hr,



followed by ligation with similarly digested vector pET-28a(+) or pACYCDuet-1 at room temperature for 1 hr. Ligation product was transformed into chemically competent *Escherichia coli* Top10 cells and screened against 50 ng/μl kanamycin for pET-28a(+) based constructs or 34 ng/μl chloramphenicol for pACYCDuet-1 based constructs. DNA insertion was confirmed through Sanger sequencing (Edinburgh Genomics, The University of Edinburgh, UK). Sequence verified constructs were transformed into *E. coli* BL21(DE3) or Tuner(DE3) for protein production. Alternatively, plasmids transformed into *E. coli* B834(DE3) cells were cultured in selenomethionine medium.

## Protein production and purification

A single colony of *E. coli* BL21(DE3) or Tuner(DE3) cells, transformed with protein expression plasmid, was transferred into 10 ml LB medium, or M9 minimal medium (MM), supplemented with appropriate antibiotic, and incubated overnight at 37 °C with 200 rpm shaking. The overnight pre-culture was then inoculated into 1 liter of LB medium and incubated at 37 °C with 200 rpm shaking. Recombinant protein production was induced at OD<sub>600</sub> = 0.6 by the addition of 1 mM IPTG and the incubation temperature was reduced to 18 °C for overnight incubation. Cells were pelleted by centrifugation at 4000 g for 20 min at 4 °C, and resuspended 10-fold (volume per gram of cell pellet) in PBS to wash cells before a second centrifugation step. Cells were resuspended in 10-times (v/w) of appropriate lysis buffer for the purification method used (see details of buffers below) and lysed by sonication on ice, with ten cycles of 30-second burst of sonication at 10 μm amplitude and 30 s of cooling. Cell lysate was clarified by centrifugation at 20,000 x g, 30 min, 4 °C; followed by filtration using a 0.22 μm syringe filter (Millipore, UK).

Selenomethionine labelled protein was produced by growing a single colony of *E. coli* B834 (DE3) cells transformed with protein expression plasmids in 100 ml LB medium supplemented with appropriate antibiotic overnight at 37 °C with shaking at 200 rpm. The overnight pre-culture was pelleted by centrifugation 3,000 x g, 4 °C, 15 min and washed twice with M9 minimal medium. The washed cells were transferred to 1 liter of SeMet medium, which contains M9 minimal medium, 40 mg/L of each L-amino acid (without methionine), 40 mg/L selenomethionine, 2 mM MgSO<sub>4</sub>, 0.4% (w/v) glucose and 1 mM Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub>. Cells were incubated at 37 °C with 200 rpm shaking and recombinant protein production was induced at OD<sub>600</sub> = 0.6 by the addition of 1 mM IPTG and the incubation temperature was reduced to 18 °C for overnight incubation. Cells were harvested and lysed as above.

## His-tagged protein purification

Clarified cell lysate was loaded onto a 5 ml HisTrap column (GE Healthcare, UK) pre-equilibrated with HisA buffer (50 mM Tris-HCl, 500 mM NaCl and 50 mM imidazole, pH 8.0). Unbound proteins were washed from the column with HisA buffer. His-tagged proteins were then eluted by a step gradient of 50% HisA buffer and 50% HisB buffer (50 mM Tris-HCl, 500 mM NaCl and 500 mM imidazole, pH 8.0). Fractions containing the protein of interest, as determined by 15% (w/v) acrylamide SDS-PAGE, were pooled before loading onto a gel-filtration column (HiLoad 16/600 Superdex 200, GE Healthcare) equilibrated with GF buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl). Fractions were subjected to 15% SDS-PAGE and those containing the protein of interest were pooled for further analysis.

## Sucrose gradient ultracentrifugation purification

Co-expressed encapsulin and EncFtn (EncFtn-Enc) and encapsulin protein were both purified according to the protocol used by M. Sutter (Sutter, 2008). Briefly, EncFtn-Enc or encapsulin was expressed based on pACYCDuet-1 vector. The *E. coli* cells were grown, induced, harvested and sonicated in a similar way as described above. GF buffer used in this purification contains 50 mM Tris-HCl, pH 8.0, and 150 mM NaCl. To remove RNA contamination, the lysate was supplemented with 50 μg/ml RNase A and rotated at 10 rpm and room temperature for 2 hrs, followed by centrifugation at 34,000 x g and 4 °C for 20 min and filtering through 0.22 μm syringe filter. Proteins were pelleted through 38% (w/v) sucrose cushion by ultracentrifugation at 100,000 x g and 4 °C for 21 hrs. 10% - 50% (w/v) sucrose gradient ultracentrifugation was applied to further separate the proteins at 100,000 x g and 4 °C for 17 hrs. Protein was dialyzed against GF buffer to remove sucrose before being used in chemical assays or TEM.

## Transmission electron microscopy

TEM imaging was performed on purified encapsulin, EncFtn, and EncFtn-Enc and apoferritin. Purified protein at 0.1 mg/ml concentration was spotted on glow-discharged 300 mesh carbon-coated copper grids and excess liquid wicked off with filter paper (Whatman, UK). The grids were washed with distilled water and blotted with filter paper three times before staining with 0.2% uranyl acetate, blotting and air-drying. Grids were imaged using a JEM1400 transmission electron microscope and images were collected with a Gatan CCD camera. Images were analyzed using ImageJ (NIH, Bethesda, MD) and size-distribution histograms were plotted using Prism 6 (GraphPad software). To observe iron mineral formation by TEM, protein samples at 8.5  $\mu$ M concentration including EncFtn<sub>sH</sub>, encapsulin, EncFtn-Enc and apoferritin were supplemented with acidic Fe(NH<sub>4</sub>)<sub>2</sub>(SO<sub>4</sub>)<sub>2</sub> at their maximum iron loading ratio in room temperature for 1 hr. The mixtures were subjected to TEM analysis with or without uranyl acetate staining. TEM experiments without Fe loading were repeated three times, a representative set of images are presented here. Proteins loaded with Fe and imaged by TEM were from single preparation.

## Protein crystallization and X-ray data collection

EncFtn<sub>sH</sub> was purified by anion exchange and Superdex 200 size-exclusion chromatography and concentrated to 10 mg/ml (based on extinction coefficient calculation). Crystallization drops were set up using the hanging drop vapor diffusion method at 292 K. Glass coverslips were set up with 1–2  $\mu$ l protein mixed with 1  $\mu$ l well solution (0.14 M calcium acetate and 15% (w/v) PEG 3350) and sealed over 1 ml of well solution. Crystals appeared after 5 days and were harvested from the well using a LithoLoop (Molecular Dimensions Limited, UK), transferred briefly to a cryoprotection solution containing well solution supplemented with 1 mM FeSO<sub>4</sub> (in 0.1% (v/v) HCl), 20% (v/v) PEG 200, and subsequently flash cooled in liquid nitrogen. Crystals of the EncFtn<sub>sH</sub> single mutations were produced in the same manner as for the EncFtn<sub>sH</sub> wild-type protein.

All crystallographic datasets were collected on the macromolecular crystallography beamlines at Diamond Light Source (Didcot, UK) at 100 K using Pilatus 6M detectors. Diffraction data were integrated and scaled using XDS (Kabsch, 2010) and symmetry related reflections were merged with Aimless (Evans, 2011). Data collection statistics are shown in Table 4. The resolution cut-off used for structure determination and refinement was determined based on the CC<sub>1/2</sub> criterion proposed by Karplus and Diederichs (2012).

The structure of EncFtn<sub>sH</sub> was determined by molecular replacement using PDB ID: 3K6C as the search model, modified to match the sequence of the target protein using Chainsaw (Stein, 2008). A single solution comprising three decamers in the asymmetric unit was found by molecular replacement using Phaser (McCoy et al., 2007). The initial model was rebuilt using Phenix.autobuild (Adams et al., 2010) followed by cycles of refinement with Phenix.refine (Afonine et al., 2012), with manual rebuilding and model inspection in Coot (Emsley et al., 2010). The final model was refined with isotropic B-factors, torsional NCS restraints, and with anomalous group refinement. The model was validated using MolProbity (Chen et al., 2010). Structural superimpositions were calculated using Coot. Crystallographic figures were generated with PyMOL. Multiple sequence alignment of EncFtn and ferritin family proteins was performed using Clustal Omega Sievers and Higgins, 2014 and displayed with Esript 3.0 (Gouet et al., 2003). Model refinement statistics are shown in Table 4. The final models and experimental data are deposited in the PDB and diffraction image files are available at the Edinburgh DataShare repository.

## Horse spleen apoferritin preparation

Horse spleen apoferritin purchased from Sigma Aldrich (UK) was dissolved in deaerated MOPS buffer (100 mM MOPS, 100 mM NaCl, 3 g/100 ml Na<sub>2</sub>S<sub>2</sub>O<sub>4</sub> and 0.5 M EDTA, pH 6.5) (Bauminger et al., 1991). Protein was dialyzed against 1 liter MOPS buffer in room temperature for two days before buffer exchanging to GF buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl) in a vivaspin column with 5 kDa cut-off (Sartorius, UK) for several times. Fe content of apoferritin was detected using ferrozine assay (Riemer et al., 2004). Protein concentration was determined using Pierce Microplate BCA Protein Assay Kit. Apoferritin containing less than 0.5 Fe per 24-mer was used in the ferroxidase assay. Apoferritin used in the Fe loading capacity experiment was prepared in the same way with 5–15 Fe per 24-mer.



## Ferroxidase assay

1 mM and 200  $\mu$ M  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$  stock solutions were prepared in 0.1% (v/v) HCl anaerobically. Protein solutions with 20  $\mu$ M FOC were diluted from ~10 mg/ml frozen stock in GF buffer (50 mM Tris-HCl, pH 8.0 and 150 mM NaCl) anaerobically. Ferroxidase activity was initiated by adding 450  $\mu$ l protein to 50  $\mu$ l of acidic  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$  at the final concentration of 100  $\mu$ M and 20  $\mu$ M in the air, respectively. The ferroxidase activity was measured by monitoring the  $\text{Fe}^{3+}$  formation which gives rise to the change of the absorbance at 315 nm (Bonomi et al., 1996). Absorbance at 315 nm was recorded every second over 1800 s using a quartz cuvette in a JASCO V-730 UV/VIS spectrophotometer (JASCO Inc., Easton, MD). In recombinantly coexpressed nanocompartments the ratio of EncFtn to Enc was assumed as 2 to 1, assuming each of the twelve pentameric vertices of the icosahedral encapsulin were occupied with decameric EncFtn. The data are presented as the mean of three technical replicates with error bars indicating one standard deviation from the mean. Proteins used here were from a single preparation.

## Iron loading capacity of ferritins

In order to determine the maximum iron loading capacity, around 8.5  $\mu$ M proteins including decameric EncFtn<sub>sH</sub>, Encapsulin, EncFtn-Enc and apoferritin were loaded with various amount of acidic  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$  ranging from 0 to 1700  $\mu$ M. Protein mixtures were incubated in room temperature for 3 hrs before desalting in Zebra spin desalting columns (7 kDa cut-off, Thermo Fisher Scientific, UK) to remove free iron ions. The protein concentration was determined using Pierce Microplate BCA assay kit (Thermo Fisher Scientific). The protein standard curve was plotted according to the manufacturer. The Fe content in the samples was determined using modified ferrozine assay (Riemer et al., 2004). Briefly speaking, 100  $\mu$ l protein sample was mixed with 100  $\mu$ l mixture of equal volume of 1.4 M HCl and 4.5% (w/v)  $\text{KMnO}_4$  and incubated at 60 °C for 2 hrs. 20  $\mu$ l of the iron-detection reagent (6.5 mM ferrozine, 6.5 mM neocuproine, 2.5 M ammonium acetate, and 1 M ascorbic acid dissolved in  $\text{H}_2\text{O}$ ) was added to the cooled tubes. 30 min later, 200  $\mu$ l of the solution was transferred into a well of 96-well plate and the absorbance at 562 nm was measured on the plate reader Spectramax M5 (Molecular Devices, UK). The standard curve was plotted using various concentrations of  $\text{FeCl}_3$  (in 10 mM HCl) diluted in the gel-filtration buffer. Three technical repeats were performed for both the ferrozine and microplate BCA assays. Samples analyzed by ICP-MS were prepared in the same way by mixing protein and ferrous ions and desalting. The proteins used in the Fe loading experiment came from a single preparation.

## Peroxidase assay

The peroxidase activity of EncFtn<sub>sH</sub> was determined by measuring the oxidation of *ortho*-phenylenediamine (OP) by  $\text{H}_2\text{O}_2$  (Pesek et al., 2011). EncFtn<sub>sH</sub> decameric and monomeric fractions purified from MM were both used in the assay. *Ortho*-phenylenediamine was prepared as a 92.5 mM stock solution in 50 mM Tris-HCl (pH 8.0). 80, 70, 60, 50, 40, 30, 20 and 10 mM of OP were prepared by diluting the stock solution in the 50 mM Tris-HCl (pH 8.0). 100  $\mu$ l of each diluted OP was added to a 96-well plate in 3 repeats. 1  $\mu$ l of 32  $\mu$ M protein was supplemented into each well to a final concentration of 160 nM, followed by the addition of 2  $\mu$ l of 30%  $\text{H}_2\text{O}_2$ . After 15 min shaking in the dark, the reaction was stopped by adding 100  $\mu$ l of 0.5 M  $\text{H}_2\text{SO}_4$ . The peroxidase activity was measured by monitoring the absorbance at 490 nm in the SpectraMax M5 Microplate Reader (Molecular Devices) (Pesek et al., 2011).

## ICP-MS analysis

Protein samples were diluted 50-fold into a solution of 2.5%  $\text{HNO}_3$  (Suprapur, Merck, UK) containing 20  $\mu$ g/L Pt as internal standard. Matrix-matched elemental standards (containing analyte metal concentrations 0 – 1000  $\mu$ g/L) were prepared by serial dilution from individual metal standard stocks (VWR) with identical solution compositions, including the internal standard. All standards and samples were analyzed by ICP-MS using a Thermo x-series instrument (Thermo Fisher Scientific) operating in collision cell mode (using 3.0 ml min<sup>-1</sup> flow of 8%  $\text{H}_2$  in He as the collision gas). Isotopes <sup>44</sup>Ca, <sup>56</sup>Fe, <sup>66</sup>Zn, <sup>78</sup>Se, and <sup>195</sup>Pt were monitored using the peak-jump method (100 sweeps, 25–30 ms dwell time on 5 channels per isotope, separated by 0.02 atomic mass units) in triplicate. The protein samples used in ICP-MS came from a single protein preparation.

## Mass spectrometry analysis

For native MS analysis, all protein samples were buffer exchanged into 100 mM ammonium acetate (pH 8.0; adjusted with dropwise addition of 1% ammonia solution) using Micro Biospin Chromatography Columns (Bio-Rad, UK) prior to analysis and the resulting protein samples were analyzed at a final concentration of ~5  $\mu$ M (oligomer concentration). In order to obtain Fe-bound EncFtn, 100  $\mu$ M or 300  $\mu$ M of freshly prepared  $\text{FeCl}_2$  was added to apo-EncFtn<sub>sH</sub> (monomer peak) immediately prior to buffer exchange into 100 mM ammonium acetate (pH 8.0). Samples were analyzed on a quadrupole ion-mobility time of flight instrument (Synapt G2, Waters Corp., Manchester, UK), equipped with a nanomate nanoelectrospray infusion robot (Advion Biosciences, Ithaca, NY). Instrument parameters were tuned to preserve non-covalent protein complexes. After optimization, typical parameters were: nanoelectrospray voltage 1.54 kV; sample cone 50 V; extractor cone 0 V; trap collision voltage 4 V; source temperature 80°C; and source backing pressure 5.5 mbar. For improved mass resolution the sample cone was raised to 155 V. Ion mobility mass spectrometry (IM-MS) was performed using the travelling-wave mobility cell in the Synapt G2, employing nitrogen as the drift gas. Typically, the IMS wave velocity was set to 300 m/s; wave height to 15 V; and the IMS pressure was 1.8 mbar. All native MS experiments were performed on samples from two independent protein preparations. For collision cross section determination, IM-MS data was calibrated using denatured equine myoglobin and data was analyzed using Driftscope v2.5 and MassLynx v4.1 (Waters Corp., UK). Theoretical collision cross sections (CCS) were calculated from pdb files using IMPACT software v. 0.9.1 (Marklund, 2015). In order to obtain information on the topology of the EncFtn<sub>sH</sub> assembly, gas-phase dissociation of the Fe-associated EncFtn<sub>sH</sub> complex was achieved by increasing the sample cone and/or trap collision voltage prior to MS analysis.

## SEC-MALLS

Size-exclusion chromatography (ÄKTA-Micro; GE Healthcare) coupled to UV, static light scattering and refractive index detection (Viscotec SEC-MALS 20 and Viscotec RI Detector:VE3580; Malvern Instruments, UK) were used to determine the molecular mass of fractions decamer and monomer of EncFtn<sub>sH</sub> in solution individually. Protein concentration was determined by measurement of absorbance at 280 nm and calculated using the extinction coefficient  $\epsilon_{0.1\%} = 1.462 \text{ mg}^{-1} \text{ ml}^{-1} \text{ cm}^{-1}$ . 100  $\mu$ l of 1.43  $\text{mg ml}^{-1}$  fractions of EncFtn<sub>sH</sub> decamer and 4.03  $\text{mg ml}^{-1}$  fractions of EncFtn<sub>sH</sub> monomer were run individually on a Superdex 200 10/300 GL size-exclusion column pre-equilibrated in 50 mM Tris-HCl (pH 8.0), 150 mM NaCl at 22°C with a flow rate of 0.5 ml/min. Light scattering, refractive index (RI) and  $A_{280\text{nm}}$  were analyzed by a homo-polymer model (OmniSEC software, v 5.1; Malvern Instruments) using the following parameters for fractions of decamer and monomer: the extinction coefficient (dA/dc) at 280 nm was 1.46 AU  $\text{mg ml}^{-1}$  and specific refractive index increment (dn/dc) was 0.185  $\text{ml g}^{-1}$ . The proteins analyzed by SEC-MALLS came from single protein preparation.

## Metal binding analysis by PAGE

Recombinant EncFtn<sub>sH</sub> fractions at 50  $\mu$ M concentration were incubated with one molar equivalent of metal ions at room temperature for 2 hrs. Half of each sample was mixed with 5 x native loading buffer (65 mM Tris-HCl, pH 8.5, 20% glycerol and 0.01% bromophenol blue) and run on non-denaturing PAGE gels (10% acrylamide) and run in Tris/glycine buffer, 200 V, 4 °C for 50 min. The remaining samples were left for an additional three hours prior to SDS-PAGE (15% acrylamide) analysis. SDS-PAGE gels were run at room temperature at 200 V, room temperature for 50 min. Gels were stained with Coomassie Brilliant Blue R250 and scanned after de-staining in water. The proteins used in this experiment came from single protein preparation.

## Analytical size-exclusion chromatography

For analysis of the multimeric state of EncFtn proteins by analytical size-exclusion gel-filtration chromatography (AGF) 25  $\mu$ l of 90  $\mu$ M protein was loaded into Superdex 200 PC 3.2/30 column (GE Healthcare) at 15 °C with GF buffer running at 0.05 ml/min and pressure limit 0.45 MPa. In order to use AGF to determine how metal ions influence the assembly of EncFtn<sub>sH</sub>, 90  $\mu$ M EncFtn<sub>sH</sub> monomer fractions were mixed with equal molar concentrations of metal ion solutions including  $\text{FeSO}_4$  in 0.1% (v/v) HCl,  $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2$ ,  $\text{FeCl}_3$ ,  $\text{CoCl}_2$ , calcium acetate (CaAc),  $\text{ZnSO}_4$  and  $\text{MnCl}_2$  at room temperature for 2 hrs prior to AGF analysis. Protein samples without metal titration were also

analyzed as a control group. Both monomer and decamer fractions of EncFtn<sub>SH</sub> left at room temperature for 2 hrs, or overnight, were also analysed as controls to show the stability of the protein samples in the absence of additional metal ions. The AGF results have been repeated twice using two independent preparations of protein, of which only one representative trace is presented in the paper.

## Accession codes and datasets

Coordinates and structure factors for the structures presented in this paper have been deposited in the PDB under the following accession codes: EncFtn<sub>SH</sub>, 5DA5; EncFtn<sub>SH</sub>-E32A, 5L89; EncFtn<sub>SH</sub>-E62A, 5L8B; EncFtn<sub>SH</sub>-H65A, 5L8G (DOIs for X-ray diffraction image data are shown in [Table 4](#)). All MS datasets presented in this paper can be found, in the raw format at [10.7488/ds/1449](https://doi.org/10.7488/ds/1449).

## Acknowledgements

This work was supported by a Royal Society Research Grant awarded to JMW (RG130585), a BBSRC New Investigator Grant to JMW and DJC (BB/N005570/1) and a Wellcome Trust Institute Strategic Support Fund grant awarded to set up the XtalPod crystallization facility at the University of Edinburgh. DJC, SH, and KA are funded by the University of Edinburgh. DH is funded by the China Scholarship Council. KJW and ET are funded by the Wellcome Trust and Royal Society through a Sir Henry Dale Fellowship awarded to KJW (098375/Z/12/Z).

We would like to thank the staff on the Macromolecular Crystallography beamlines at Diamond Light Source for their assistance with data collection. We would like to thank Dr Steve Mitchell in the University of Edinburgh and Electron Microscopy Research Services in Newcastle University for the use of Transmission Electron Microscopes. We would like to thank Dr Michael Capeness and Dr. Louise Horsfall for the use of their anaerobic chamber. We would like to thank staff in the Edinburgh Protein Production Facility for their guidance and patience, Dr Janice Brahman, Dr Liz Blackburn, Dr Matthew Nowicki and Dr Martin Wear. We would like to thank Professor Rick Lewis and Dr Arnaud Basle (Newcastle University, UK), and Dr John Berrisford (PDBe, EMBL-EBI, Hinxton, Cambridge) for help identifying the glycolic acid ligand in the ferroxidase center. We would like to thank Prof Dominic Campopiano, Dr Elisabeth Lowe and Laura Tuck for their critical reading of this manuscript and helpful discussions.

## Additional information

### Funding

Funder	Grant reference number	Author
China Scholarship Council		Didi He
University Of Edinburgh		Sam Hughes Kirsten Altenbach David J Clarke
Wellcome Trust	098375/Z/12/Z	Emma Tarrant Kevin J Waldron
Biotechnology and Biological Sciences Research Council	BB/N005570/1	David J Clarke Jon Marles-Wright
Royal Society	RG130585	Jon Marles-Wright

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

DH, DJC, JM-W, Conception and design, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article; SH, SV-H, ET, Acquisition of data, Analysis and interpretation of data; AG, Acquisition of data, Drafting or revising the article; KA, CLM, Acquisition of data, Contributed unpublished essential data or reagents; KJW, Acquisition of data, Analysis and interpretation of data, Drafting or revising the article

## Author ORCIDs

Didi He, <http://orcid.org/0000-0002-3360-9352>

Kevin J Waldron, <http://orcid.org/0000-0002-5577-7357>

David J Clarke, <http://orcid.org/0000-0002-3741-2952>

Jon Marles-Wright, <http://orcid.org/0000-0002-9156-3284>

## Additional files

## Supplementary files

• Supplementary file 1. Multiple sequence alignment of ferritin-like proteins. The amino acid sequences of ferritin family proteins were aligned progressively using EMBL-EBI web services (Sutter et al., 2008; Akita et al., 2007), Clustal Omega (McHugh et al., 2014) T-Coffee (Contreras et al., 2014) and MAFFT (Helgstrand et al., 2003). Protein names were adapted from either UniprotKB (Sutter et al., 2008), KEGG (Roberts et al., 2011) database or common name with PDB entry code (He and Marles-Wright, 2015). Sequences were sorted in an order corresponding to the clades in phylogenetic tree (Figure 13). The alignment was edited by Esprout 3.0 web server (Aberg et al., 1993). The *R. rubrum* EncFtn (Rru\_A0973) sequence is highlighted in yellow. The ferroxidase centre (FOC) of *Pseudo-nitzschia multiseriata* ferritin (PmFtn\_4ITW) (highlighting in blue) consists of Fe<sub>A</sub> site (E16, E49, E52) and Fe<sub>B</sub> site (E49, E95, E131, Q128) which are labelled with solid red triangles (Grant et al., 1998). Another iron binding site in PmFtn\_4ITW (the gateway site or Fe<sub>C</sub> site [Bradley et al., 2014]) consists of E48, E45 and E131 which are marked with solid blue circles (Grant et al., 1998). The FOC of *R. rubrum* EncFtn is labelled with empty red triangles as E32, E62, H65 and Y39; and the iron entry site is marked with empty blue circles including E31 and E34. The putative iron exit site is marked with empty blue squares including H57, E61 and E64. The C-terminal localization sequences common to the encapsulin associated ferritins are highlighted within the red rectangle.

DOI: [10.7554/eLife.18972.035](https://doi.org/10.7554/eLife.18972.035)

• Supplementary file 2. Primers used in this study. Primers used to generate the original constructs used in this study are listed 5' to 3', from left to right. Introduced restriction sites are shown underlined; regions complementary to genomic DNA shown in bold.

DOI: [10.7554/eLife.18972.036](https://doi.org/10.7554/eLife.18972.036)

## Major datasets

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database, license, and accessibility information
Hughes S, Vanden-Hehir S, Clarke D	2016	Native MS Analysis of Encapsulated Ferritin from <i>Rhodospirillum rubrum</i>	<a href="http://dx.doi.org/10.7488/ds/1449">http://dx.doi.org/10.7488/ds/1449</a>	Publicly available at Edinburgh DataShare
He D, Georgiev A, Marles-Wright J	2016	Single crystal X-ray diffraction data for <i>Rhodospirillum rubrum</i> encapsulated ferritin Rru_A0973, short His-tagged variant	<a href="http://dx.doi.org/10.7488/ds/1342">http://dx.doi.org/10.7488/ds/1342</a>	Publicly available at Edinburgh DataShare
He D, Marles-Wright J	2016	Single crystal X-ray diffraction data for <i>Rhodospirillum rubrum</i> encapsulated ferritin Rru_A0973, short His-tagged variant E32A mutant	<a href="http://dx.doi.org/10.7488/ds/1419">http://dx.doi.org/10.7488/ds/1419</a>	Publicly available at Edinburgh DataShare
He D, Marles-Wright J	2016	Single crystal X-ray diffraction data for <i>Rhodospirillum rubrum</i> encapsulated ferritin Rru_A0973, short His-tagged variant E62A mutant	<a href="http://dx.doi.org/10.7488/ds/1420">http://dx.doi.org/10.7488/ds/1420</a>	Publicly available at Edinburgh DataShare
He D, Marles-Wright J	2016	Single crystal X-ray diffraction data for <i>Rhodospirillum rubrum</i> encapsulated ferritin Rru_A0973, short His-tagged variant H65A mutant	<a href="http://dx.doi.org/10.7488/ds/1421">http://dx.doi.org/10.7488/ds/1421</a>	Publicly available at Edinburgh DataShare

## References

- Aberg A**, Nordlund P, Eklund H. 1993. Unusual clustering of carboxyl side chains in the core of iron-free ribonucleotide reductase. *Nature* **361**:276–278. doi: [10.1038/361276a0](https://doi.org/10.1038/361276a0)
- Adams PD**, Afonine PV, Bunkóczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ, Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson JS, Terwilliger TC, Zwart PH. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**:213–221. doi: [10.1107/S0907444909052925](https://doi.org/10.1107/S0907444909052925)
- Afonine PV**, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. 2012. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D Biological Crystallography* **68**:352–367. doi: [10.1107/S0907444912001308](https://doi.org/10.1107/S0907444912001308)
- Akita F**, Chong KT, Tanaka H, Yamashita E, Miyazaki N, Nakaishi Y, Suzuki M, Namba K, Ono Y, Tsukihara T, Nakagawa A. 2007. The crystal structure of a virus-like particle from the hyperthermophilic archaeon *Pyrococcus furiosus* provides insight into the evolution of viruses. *Journal of Molecular Biology* **368**:1469–1483. doi: [10.1016/j.jmb.2007.02.075](https://doi.org/10.1016/j.jmb.2007.02.075)
- Andrews SC**. 2010. The Ferritin-like superfamily: Evolution of the biological iron storeman from a rubrerythrin-like ancestor. *Biochimica Et Biophysica Acta (BBA) - General Subjects* **1800**:691–705. doi: [10.1016/j.bbagen.2010.05.010](https://doi.org/10.1016/j.bbagen.2010.05.010)
- Bauminger ER**, Harrison PM, Hechel D, Nowik I, Treffry A. 1991. Mössbauer spectroscopic investigation of structure-function relations in ferritins. *Biochimica Et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology* **1118**:48–58. doi: [10.1016/0167-4838\(91\)90440-B](https://doi.org/10.1016/0167-4838(91)90440-B)
- Behera RK**, Theil EC. 2014. Moving Fe<sup>2+</sup> from ferritin ion channels to catalytic OH centers depends on conserved protein cage carboxylates. *PNAS* **111**:7925–7930. doi: [10.1073/pnas.1318417111](https://doi.org/10.1073/pnas.1318417111)
- Bertini I**, Lalli D, Mangani S, Pozzi C, Rosa C, Theil EC, Turano P. 2012. Structural insights into the ferroxidase site of ferritins from higher eukaryotes. *Journal of the American Chemical Society* **134**:6169–6176. doi: [10.1021/ja210084n](https://doi.org/10.1021/ja210084n)
- Beveridge R**, Covill S, Pacholarz KJ, Kalapothakis JMD, MacPhee CE, Barran PE. 2014. A Mass-Spectrometry-Based Framework To Define the Extent of Disorder in Proteins. *Analytical Chemistry* **86**:10979–10991. doi: [10.1021/ac5027435](https://doi.org/10.1021/ac5027435)
- Bonomi F**, Kurtz Jr. DM, Cui X. 1996. Ferroxidase activity of recombinant *Desulfovibrio vulgaris* rubrerythrin. *Journal of Biological Inorganic Chemistry* **1**:67–72. doi: [10.1007/s007750050024](https://doi.org/10.1007/s007750050024)
- Bou-Abdallah F**. 2010. The iron redox and hydrolysis chemistry of the ferritins. *Biochimica Et Biophysica Acta (BBA) - General Subjects* **1800**:719–731. doi: [10.1016/j.bbagen.2010.03.021](https://doi.org/10.1016/j.bbagen.2010.03.021)
- Bradley JM**, Moore GR, Le Brun NE. 2014. Mechanisms of iron mineralization in ferritins: one size does not fit all. *JBIC Journal of Biological Inorganic Chemistry* **19**:775–785. doi: [10.1007/s00775-014-1136-3](https://doi.org/10.1007/s00775-014-1136-3)
- Cassidy-Amstutz C**, Oltrogge L, Going CC, Lee A, Teng P, Quintanilla D, East-Seletsky A, Williams ER, Savage DF. 2016. Identification of a Minimal Peptide Tag for in Vivo and in Vitro Loading of Encapsulin. *Biochemistry* **55**:3461–3468. doi: [10.1021/acs.biochem.6b00294](https://doi.org/10.1021/acs.biochem.6b00294)
- Chasteen ND**, Harrison PM. 1999. Mineralization in ferritin: an efficient means of iron storage. *Journal of Structural Biology* **126**:182–194. doi: [10.1006/jsbi.1999.4118](https://doi.org/10.1006/jsbi.1999.4118)
- Chen VB**, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography* **66**:12–21. doi: [10.1107/S0907444909042073](https://doi.org/10.1107/S0907444909042073)
- Contreras H**, Joens MS, McMath LM, Le VP, Tullius MV, Kimmey JM, Bionghi N, Horwitz MA, Fitzpatrick JA, Goulding CW. 2014. Characterization of a *Mycobacterium tuberculosis* nanocompartment and its potential cargo proteins. *Journal of Biological Chemistry* **289**:18279–18289. doi: [10.1074/jbc.M114.570119](https://doi.org/10.1074/jbc.M114.570119)
- Emsley P**, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography* **66**:486–501. doi: [10.1107/S0907444910007493](https://doi.org/10.1107/S0907444910007493)
- Evans PR**. 2011. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D Biological Crystallography* **67**:282–292. doi: [10.1107/S090744491003982X](https://doi.org/10.1107/S090744491003982X)
- Fernandez de la Mora J**. 2000. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta* **406**:93–104. doi: [10.1016/S0003-2670\(99\)00601-7](https://doi.org/10.1016/S0003-2670(99)00601-7)
- Folta-Stogniew E**. 2006. Oligomeric states of proteins determined by size-exclusion chromatography coupled with light scattering, absorbance, and refractive index detectors. *Methods in Molecular Biology* **328**:97–112. doi: [10.1385/1-59745-026-X:97](https://doi.org/10.1385/1-59745-026-X:97)
- Gouet P**, Robert X, Courcelle E. 2003. ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research* **31**:3320–3323. doi: [10.1093/nar/gkg556](https://doi.org/10.1093/nar/gkg556)
- Grant RA**, Filman DJ, Finkel SE, Kolter R, Hogle JM. 1998. The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nature Structural Biology* **5**:294–303. doi: [10.1038/nsb0498-294](https://doi.org/10.1038/nsb0498-294)
- Haldar S**, Bevers LE, Tosha T, Theil EC. 2011. Moving Iron through ferritin protein nanocages depends on residues throughout each four  $\alpha$ -helix bundle subunit. *Journal of Biological Chemistry* **286**:25620–25627. doi: [10.1074/jbc.M110.205278](https://doi.org/10.1074/jbc.M110.205278)
- Hall Z**, Hernández H, Marsh JA, Teichmann SA, Robinson CV. 2013. The Role of Salt Bridges, Charge Density, and Subunit Flexibility in Determining Disassembly Routes of Protein Complexes. *Structure* **21**:1325–1337. doi: [10.1016/j.str.2013.06.004](https://doi.org/10.1016/j.str.2013.06.004)



- He D, Marles-Wright J. 2015. Ferritin family proteins and their use in bionanotechnology. *New Biotechnology* **32**: 651–657. doi: [10.1016/j.nbt.2014.12.006](https://doi.org/10.1016/j.nbt.2014.12.006)
- Helgstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L. 2003. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. *Journal of Molecular Biology* **334**:885–899. doi: [10.1016/j.jmb.2003.09.035](https://doi.org/10.1016/j.jmb.2003.09.035)
- Honarmand Ebrahimi K, Bill E, Hagedoorn PL, Hagen WR. 2012. The catalytic center of ferritin regulates iron storage via Fe(II)-Fe(III) displacement. *Nature Chemical Biology* **8**:941–948. doi: [10.1038/nchembio.1071](https://doi.org/10.1038/nchembio.1071)
- Kabsch W. 2010. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D Biological Crystallography* **66**:133–144. doi: [10.1107/S0907444909047374](https://doi.org/10.1107/S0907444909047374)
- Karplus PA, Diederichs K. 2012. Linking crystallographic model and data quality. *Science* **336**:1030–1033. doi: [10.1126/science.1218231](https://doi.org/10.1126/science.1218231)
- Katz AK, Glusker JP, Beebe SA, Bock CW. 1996. Calcium Ion Coordination: A Comparison with That of Beryllium, Magnesium, and Zinc. *Journal of the American Chemical Society* **118**:5752–5763. doi: [10.1021/ja953943i](https://doi.org/10.1021/ja953943i)
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* **33**:1870–1874. doi: [10.1093/molbev/msw054](https://doi.org/10.1093/molbev/msw054)
- Lawson DM, Artymiuk PJ, Yewdall SJ, Smith JM, Livingstone JC, Treffry A, Luzzago A, Levi S, Arosio P, Cesareni G. 1991. Solving the structure of human H ferritin by genetically engineering intermolecular crystal contacts. *Nature* **349**:541–544. doi: [10.1038/349541a0](https://doi.org/10.1038/349541a0)
- Le Brun NE, Crow A, Murphy MEP, Mauk AG, Moore GR. 2010. Iron core mineralisation in prokaryotic ferritins. *Biochimica Et Biophysica Acta (BBA) - General Subjects* **1800**:732–744. doi: [10.1016/j.bbagen.2010.04.002](https://doi.org/10.1016/j.bbagen.2010.04.002)
- Lundin D, Poole AM, Sjöberg BM, Högbom M. 2012. Use of structural phylogenetic networks for classification of the ferritin-like superfamily. *Journal of Biological Chemistry* **287**:20565–20575. doi: [10.1074/jbc.M112.367458](https://doi.org/10.1074/jbc.M112.367458)
- Mann S, Bannister JV, Williams RJ. 1986. Structure and composition of ferritin cores isolated from human spleen, limpet (*Patella vulgata*) hemolymph and bacterial (*Pseudomonas aeruginosa*) cells. *Journal of Molecular Biology* **188**:225–232. doi: [10.1016/0022-2836\(86\)90307-4](https://doi.org/10.1016/0022-2836(86)90307-4)
- Marklund EG, Degiacomi MT, Robinson CV, Baldwin AJ, Benesch JL. 2015. Collision cross sections for structural collision cross sections for structural proteomics. *Structure* **23**:791–799. doi: [10.1016/j.str.2015.02.010](https://doi.org/10.1016/j.str.2015.02.010)
- McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. *Journal of Applied Crystallography* **40**:658–674. doi: [10.1107/S0021889807021206](https://doi.org/10.1107/S0021889807021206)
- McHugh CA, Fontana J, Nemecek D, Cheng N, Aksyuk AA, Heymann JB, Winkler DC, Lam AS, Wall JS, Steven AC, Hoiczky E. 2014. A virus capsid-like nanocompartment that stores iron and protects bacteria from oxidative stress. *The EMBO Journal* **33**: e201488566. doi: [10.15252/embj.201488566](https://doi.org/10.15252/embj.201488566)
- Nei M, Kumar S. 2000. *Molecular Evolution and Phylogenetics*. Oxford university press.
- Pesek J, Büchler R, Albrecht R, Boland W, Zeth K. 2011. Structure and mechanism of iron translocation by a Dps protein from *Microbacterium arborescens*. *Journal of Biological Chemistry* **286**:34872–34882. doi: [10.1074/jbc.M111.246108](https://doi.org/10.1074/jbc.M111.246108)
- Pfaffen S, Abdulqadir R, Le Brun NE, Murphy ME. 2013. Mechanism of ferrous iron binding and oxidation by ferritin from a pennate diatom. *Journal of Biological Chemistry* **288**:14917–14925. doi: [10.1074/jbc.M113.454496](https://doi.org/10.1074/jbc.M113.454496)
- Pope MR, Murrell SA, Ludden PW. 1985. Covalent modification of the iron protein of nitrogenase from *Rhodospirillum rubrum* by adenosine diphosphoribosylation of a specific arginine residue. *Proceedings of the National Academy of Sciences* **82**:3173–3177. doi: [10.1073/pnas.82.10.3173](https://doi.org/10.1073/pnas.82.10.3173)
- Riemer J, Hoepken HH, Czerwinska H, Robinson SR, Dringen R. 2004. Colorimetric ferrozine-based assay for the quantitation of iron in cultured cells. *Analytical Biochemistry* **331**:370–375. doi: [10.1016/j.ab.2004.03.049](https://doi.org/10.1016/j.ab.2004.03.049)
- Roberts JN, Singh R, Grigg JC, Murphy ME, Bugg TD, Eltis LD. 2011. Characterization of dye-decolorizing peroxidases from *Rhodococcus jostii* RHA1. *Biochemistry* **50**:5108–5119. doi: [10.1021/bi200427h](https://doi.org/10.1021/bi200427h)
- Rurup WF, Cornelissen JJ, Koay MS. 2015. Recombinant expression and purification of "virus-like" bacterial encapsulin protein cages. *Methods in Molecular Biology* **1252**:61–67. doi: [10.1007/978-1-4939-2131-7\\_6](https://doi.org/10.1007/978-1-4939-2131-7_6)
- Rurup WF, Snijder J, Koay MS, Heck AJ, Cornelissen JJ. 2014. Self-sorting of foreign proteins in a bacterial nanocompartment. *Journal of the American Chemical Society* **136**:3828–3832. doi: [10.1021/ja410891c](https://doi.org/10.1021/ja410891c)
- Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**:406–425.
- Sievers F, Higgins DG. 2014. Clustal Omega, accurate alignment of very large numbers of sequences. *Methods in Molecular Biology* **1079**:105–116. doi: [10.1007/978-1-62703-646-7\\_6](https://doi.org/10.1007/978-1-62703-646-7_6)
- Stein N. 2008. CHAINSAW : a program for mutating pdb files used as templates in molecular replacement. *Journal of Applied Crystallography* **41**:641–643. doi: [10.1107/S0021889808006985](https://doi.org/10.1107/S0021889808006985)
- Sutter M, Boehringer D, Gutmann S, Günther S, Prangishvili D, Loessner MJ, Stetter KO, Weber-Ban E, Ban N. 2008. Structural basis of enzyme encapsulation into a bacterial nanocompartment. *Nature Structural & Molecular Biology* **15**:939–947. doi: [10.1038/nsmb.1473](https://doi.org/10.1038/nsmb.1473)